Fisheries and Oceans Canada | Pêches et Océans Canada
Science | Sciences

C S A S | **S C C S**

Canadian Science Advisory Secretariat | **Secrétariat canadien de consultation scientifique**

Research Document 2002/121 | **Document de recherche 2002/121**

Not to be cited without permission of the authors * | Ne pas citer sans autorisation des auteurs *

Review of data quality and sampling protocols of the Shorekeepers' Guide

Examen de la qualité des données et des protocoles d'échantillonnage du Guide des gardiens du littoral

G.S. Jamieson[1], R. Scrosati[2], C. Schwarz[3], C. Levings[4], and B. Smiley[5]

[1]Department of Fisheries and Oceans
Pacific Biological Station, Nanaimo, B.C. V9T 6N7

[2]University of British Columbia, Department of Botany
Vancouver, B.C. V6T 1Z4

[3] Department of Statistics and Actuarial Science
Simon Fraser University, 8888 University Drive
Burnaby, B.C. V5A 1S6

[4]Department of Fisheries and Oceans
West Vancouver Laboratory, 4160 Marine Dr.
West Vancouver, B.C. V7V1N6

[5] Department of Fisheries and Oceans
Institute of Ocean Science, Sidney, BC V8L 4B2

* This series documents the scientific basis for the evaluation of fisheries resources in Canada. As such, it addresses the issues of the day in the time frames required and the documents it contains are not intended as definitive statements on the subjects addressed but rather as progress reports on ongoing investigations.

Research documents are produced in the official language in which they are provided to the Secretariat.

This document is available on the Internet at:

* La présente série documente les bases scientifiques des évaluations des ressources halieutiques du Canada. Elle traite des problèmes courants selon les échéanciers dictés. Les documents qu'elle contient ne doivent pas être considérés comme des énoncés définitifs sur les sujets traités, mais plutôt comme des rapports d'étape sur les études en cours.

Les documents de recherche sont publiés dans la langue officielle utilisée dans le manuscrit envoyé au Secrétariat.

Ce document est disponible sur l'Internet à:

http://www.dfo-mpo.gc.ca/csas/

ISSN 1480-4883

© Her Majesty the Queen in Right of Canada, 2002
© Sa majesté la Reine, Chef du Canada, 2002

Canada

Abstract

The Shorekeepers' Guide, released by DFO in 1999 following three years of development, is currently being used by a number of volunteer groups through out coastal British Columbia in the monitoring of intertidal ecosystems. Data have been collected since 1997, and while improvements in Guide protocols continue to be made, it requires a multi-year data series to begin a process of data quality and analysis evaluation. Here, we present results from analyses of data that have been collected at the same locations over time to evaluate the accuracy and precision with which data are being reported, and the utility of existing data recording procedures. Our analyses need to be considered in the context that 1) some of the data analysed here were collected early in the program's development, i.e., some problems described here were identified in other ways and have already been dealt with, and 2) in the single year audit of sampling procedures reported here, logistic difficulties resulted in an excessively long time period between samplings, with the result that in some site audits, seasonal differences in community structure made the detection of possible data collection inconsistencies impossible. Nevertheless, many of the observations and recommendations presented here are relevant and constructive. Recommendations have been or are being incorporated into on-going survey and analytical procedures, and past shorekeeper data are being edited, where possible, so as to ensure the most accurate and credible database exists.

Résumé

Le Guide des gardiens du littoral, publié par le MPO en 1999 après trois années de préparation, est présentement utilisé par divers groupes de bénévoles à l'échelle de la région côtière de la Colombie-Britannique pour surveiller l'état des écosystèmes intertidaux. Des données ont été recueillies depuis 1997 et, même si on continue à peaufiner les protocoles du Guide, une série de données s'échelonnant sur plusieurs années est requise pour évaluer leur qualité et les analyser. Sont présentés les résultats d'analyses de données recueillies aux mêmes endroits au fil du temps faites en vue d'évaluer la précision et l'exactitude des données consignées et l'utilité des procédures actuelles d'enregistrement des données. ces analyses doivent être considérées à la lumière des deux facteurs suivants : 1) certaines des données analysées ont été recueillies au début du programme c'est-à-dire que certains des problèmes décrits ont été identifiés par d'autres moyens et ont déjà été réglés et 2) lors de la vérification des procédures d'échantillonnage dont il est question dans le présent rapport, des problèmes logistiques ont résulté en un écart excessivement long entre les échantillonnages, ce qui a fait que, dans le cas de la vérification des données sur certains sites, il a été impossible de détecter le manque de cohérence dans les données recueillies à cause de différences saisonnières dans la structure des communautés. Malgré cela, nombre des observations et des recommandations présentées sont pertinentes et constructives. Les recommandations ont été incorporées aux procédures de relevé et d'analyse, ou sont en voie de l'être, et les données recueillies par les gardiens par le passé sont en voie d'être épurées, lorsque possible, de sorte à assurer que la base de données ainsi établie soit aussi précise et fiable que possible.

Abstract 2
Résumé 2
1 Introduction 4
2 Results of Data Quality Evaluation 4
2.1 Utilisation of sampling procedures described in the guide 5
2.1.1 Objective: 5
2.1.2 Methods: 5
2.1.3 Results: 5
2.1.4 Recommendations 7
2.2 Degree of spatial consistency 8
2.2.1 Objective: 8
2.2.2 Methods: 8
2.2.3 Results 8
2.2.4 Discussion 10
2.3 Accuracy of identifications of plants and animals 11
2.3.1 Objective: 11
2.3.2 Analysis: 11
2.3.3 Recommendations 11
2.4 Transcription error rate between field sheets and computer spreadsheets 12
2.4.1 Objective: 12
2.4.2 Methods: 12
2.4.3 Recommendations 13
2.5 Field audits 14
2.5.1 Objective: 14
2.5.2 Methods: 14
2.5.3 Results and Discussion 14
3 Survey Design and Data Analysis Evaluation 16
3.1 Design Issues 16
3.1.2 Permanent monitoring sites 17
3.1.3 Laying out transects and quadrats: 19
3.1.4 Impact of repeated sampling over time 19
3.2 Data collection and analysis 20
3.2.1 Large counts: 20
3.2.2 Depth of excavation 20
3.2.3 Derived variables 20
3.2.4 Missing data/zeroes 20
3.3 Analyses 21
3.3.1 Changes in mean abundance 21
3.3.2 Species composition 22
3.4 Recommendations from a data analysis perspective 23
4 Discussion 23
5 Recommendations 24
6 Acknowledgements 24
7 References 25

1 Introduction

In 1996, contractors working under our direction prepared a draft manual with a number of modules to provide practical field tools for groups of citizens interested in obtaining scientifically defensible monitoring data on intertidal ecosystems and habitats (Smiley and Levings 1996). This work was in support of implementation of the Canada Oceans Act. The manual was of considerable interest to habitat and ocean managers and the general public. The modules were refined and rewritten in subsequent years, culminating in 1999 with a publication entitled the "Shorekeepers' Guide" (Jamieson et al. 1999). The intent of the Shorekeepers Program was to utilise community volunteers under the direction of a trained leader to collect intertidal biological data relevant to ecosystem-based management and marine environmental quality (MEQ) monitoring.

We recognised that methodology would evolve as experience was gained in using the Guide (Jamieson et al 1999). Between 1996 and 2000, a variety of community groups were engaged in data collection, mainly on the east coast of Vancouver Island. Between 1996 and 1998 there were three major field testings of the methods and materials, with each leading to significant improvements or adjustments in the methodology. A few minor modifications were made after 1999.

By 2000 a large amount of data had been collected and archived at the Pacific Biological Station in data bases presently maintained by Mr. T.J. Brown, but because of time constraints and priorities, they had not been examined in detail or synthesised. In October 2000, we were approached by Dr. John Pringle, Division Head, MEHSD, with a request to review the results of the program to date. As part of the ongoing review process, an "audit" of sampling procedures was initiated in 2001 to assess the quality of data being obtained, the types of problems that were being encountered in data acquisition, and to make recommendations as to what further improvements might be made in either the field protocols being used, data analysis or leader training. Funding from the Environmental Sciences Strategic Research Fund (ESSRF) was obtained to contract outside expertise in this evaluation, which was undertaken by Dr. Ricardo Scrosati, a UBC biologist, and Dr. Carl Schwarz, a SFU statistician. This document is a merging of their reports, which have been edited to remove duplication, clarify ambiguous statements, and add additional relevant information. The specific objective was "to test the null hypothesis that data obtained by volunteer groups on the distribution of intertidal habitat types and abundance of organisms within habitats do not meet scientific standards".

The main objective of the Guide is the monitoring of biological intertidal communities along the coast of British Columbia. From 1997 onwards, a number of local biologists, from south-east Vancouver Island in particular, were trained in basic ecological methodology to prepare them to lead the surveys. Guidelines to perform the surveys are compiled in the Shorekeepers guide.

While the leaders were trained biologists, most of the people working under them who did much of the work in the surveys were not professional marine biologists, so here we assess the degree of reliability of their data. Leaders are tasked in the Guide with ensuring that data quality (e.g., species identification, quadrate placement, etc.) is acceptable, and that data are recorded appropriately. This audit is thus intended in part to assess the extent to which data collection and monitoring quality was maintained, and to address any issues that might arise if this was not the case.

2 Results of Data Quality Evaluation

Data sets produced by the volunteer groups between 1997 and 1999 on south-eastern Vancouver Island were examined. While this is the only multi-year data comparable at this time, in hind-sight they are perhaps not the most appropriate data to be used in an analysis of this type for the following reasons:

1) The protocol was still being developed, with minor changes being made as experience was being acquired in using it under field conditions;

2) Funding for sampling was obtained from a variety of sources, and each source typically involved new participants (e.g., displaced fishers), not all of whom perhaps had the same motivation or skills to report data accurately, and whom would have been likely to have made perhaps proportionally more errors in their early efforts; and

3) We had not yet perfected our leader training program, so even the leaders to some extent were learning what to do while actually conducting surveys, which may have introduced more data errors than would be expected from leaders conducting surveys today.

Original data have been archived with Mr. Tom J. Brown, Pacific Biological Station, Nanaimo, British Columbia.

The review emphasised the following five topics: utilisation of sampling procedures described in the guide, degree of spatial consistency, accuracy of identifications of plants and animals, transcription error rate between field sheets and computer spreadsheets, and field audits.

2.1 Utilisation of sampling procedures described in the guide

2.1.1 Objective:

To estimate the degree of departure of survey teams from the sampling procedures described in the Shorekeepers guide.

2.1.2 Methods:

The field sheets produced by survey teams were compared with the Shorekeepers guide (Jamieson et al. 1999) to assess whether they fully followed the expected sequential operational steps outlined in the Shorekeepers' Guide.

2.1.3 Results:

Between 1997 and 1999, a total of 109 surveys were completed (33 in 1997, 28 in 1998, and 48 in 1999) at 52 different sites. Table 1 gives the list of surveyed sites, their respective Shorekeepers codes, survey dates, and the code letter for the survey leader. Team volunteers varied among surveys and not always specified in full in the appropriate place in field sheets, so comparisons among surveys based on team members were difficult.

In general, survey teams followed all of the steps (total = 34 steps) described in the Shorekeepers guide when doing their surveys. However, there were a number of problems in the way in which each quantitative step was done or reported. Below, we provide an account of such problems. Several of the 34 steps were of organisational nature (*e.g.* Step 9: Assembly of survey team and materials), so they could not be analysed quantitatively. Thus, only the quantitative steps are discussed below.

2.1.3.1 Step 13: Identification of habitats

Habitat classification and area determination is by definition somewhat subjective, which was why the Guide recommends that it be a collective consensus by all members of a survey team. However, real errors can still occur, as discussed below. At least one of the observed habitats was not identified as specified by the Guide in 59% of the surveys, with little differences between years (Table 2). Problems were:

• Identification of more than one patch [*e.g.* using C1, C2, etc. for habitat C (cobble)] for a habitat type that was continuous in space (28% of the surveys). Only when two or more spatially separated patches of the same habitat type are found should patches be quantified separately.
• Incorrect naming of habitat types (23% of the surveys), without properly taking into account the % cover of species that define habitat types. For example, sometimes, a habitat type was incorrectly named as FA1 (name reserved for habitats with 50%, or more, cover of *Fucus*) when *Fucus* covered less than 50%. For those cases, the habitat should have been named OA1 (other algae), if the total % cover of seaweeds was

50% or more, or C1 (cobble), S1 (sand), R1 (rock), or U1 (mud), depending on the primary substrate, if algae covered less than 50% of the patch.

- In 8% of the surveys, habitat patches were identified in lists but not named in the sketch map, although all of the patches had been drawn.
- In 5% of the surveys, different patches of the same habitat type were correctly identified and named in the map, but a lower number of patches was reported in the final list, thus making it difficult to know which patches were quantified.
- 1% of the surveys described a number of patches of a variety of habitat types in lists, but without mapping them at all (not even their outlines on the sketch map).

2.1.3.2 Step 15: Location of baseline end points

There were problems in the way of establishing, or noting, the baseline end points in 19% of the surveys, with little differences between years (Table 2). Problems were:

- In 14% of the surveys, the distance between the baseline end points and the benchmarks were not noted, so it was likely impossible to re-locate the baseline exactly in the same place in successive surveys.
- In 6% of the surveys, some of the distances between the same benchmarks and the baseline end points differed from the previous survey by a few meters, which likely resulted in laying the baseline in slightly different places.
- In 2% of the surveys, new benchmarks that differed from the original ones were established, which likely did not allow placement of the baseline in the exact same place in successive surveys.
- Many of the electronic maps (obtained by scanning the original sketch maps) were of poor quality, which did not allow the distance between benchmarks and the baseline end points to be read clearly. Although the original maps are available for double-checking, the aim of having electronic maps is to work with them and to exchange them among colleagues, so it is necessary that electronic maps be of good quality.

2.1.3.3 Step 18: Map scales

The scale that applied to the width (perpendicular to the shore line) and length (parallel to the shore line) of the study area was not always the same, because it depended on the width of the intertidal zone on the survey date (a relatively random factor) and on the length chosen by survey teams (a fixed, arbitrary factor). Both scales were generally noted appropriately in maps. In 6% of the surveys, scaling presented problems (Table 2), normally simply due to the lack of noting the length scale in the map. The width was occasionally entered as the baseline length; data from 1997 to present have been assessed for width = 50 or 100 and checked against the mapping, with corrections then made where necessary.

2.1.3.4 Step 21: Geographic co-ordinates

Geographic co-ordinates were incorrectly determined for 5% of the surveys (Table 2). Two values corresponded to places that were well inland, whereas the other three corresponded to places located in deep water.

2.1.3.5 Step 22: Habitat area

Determination of habitat area was incorrectly done for one or more habitats in 16% of the surveys, with error values lower than 20% for each of the three years analysed (Table 2). The common problem was that area estimations, either for an entire habitat or for some of its patches, did not always correspond to what could be calculated using the corresponding map. Since areas are calculated from the maps, it is assumed that calculations were done incorrectly. There is no way to evaluate *a posteriori* whether the maps were drawn correctly.

2.1.3.6 Step 23: Habitat slope

Slopes were generally well determined, with wrong slopes recorded for just one habitat from each of only two surveys out of the 109 surveys. '0' slopes were reported for habitats that had a clear difference in minimum and maximum elevations. The error rate for slope determinations was thus **1.8%**. A recorded slope with a value "99" indicates no slope was measured; a slope has to be entered before further data could be entered on the spread sheet, and 99 was the default value. Usually a note or comment is attached to the habitat category explaining why this occurred.

2.1.3.7 Step 24: Habitat elevation

Determination of habitat elevation was reported incorrectly for one or more habitats in 37% of the surveys, without much interannual variation (Table 2). For two contiguous habitats, sometimes the maximum elevation of the lower habitat was lower than the minimum elevation of the higher habitat, leaving a gap in elevation that was physically impossible to have occurred. Additionally, a few elevation measurements were much higher than 4 m (the approximate maximum level of high tide for the Strait of Georgia). Sometimes, the MIN and MAX elevation are reversed. Again, values here have to be entered before further data can be entered into the spreadsheet.

2.1.3.8 Step 27: Location of transects and quadrats

Explicit location of either transects and quadrats in the map is a requirement of the Guide's protocol. The Guide states that quadrat locations within a habitat should be systematically placed along transects that divide the habitat's dimensions into 3-4 equal units. An ideal sampling design (Krebs 1999) should have a random arrangement of quadrats (so they will likely fall in different spots in successive surveys), but a systematic one was used as it is simpler to communicate to lay people (see survey design discussion below). Given this, defining explicit quadrat locations is acceptable; with no protocol for quadrat locating specified, there is likely to be bias in site selection.

2.1.3.9 Step 31: Estimation of plant/animal abundance

The final computer-generated reports that can be produced from the database give lists of abundance (% cover or density of individuals) of plants and animals observed in the field. However, abundance was not specified for all the identified taxa in **77%** of the reports, with a highest value (94%) in 1997 (Table 2). These values apply to computer tables derived from field sheets. These tables are the key basic data for statistical analyses, which constitute the base of further analyses.

- The average percentage of taxa that were listed, but for which no abundance measures were provided, was 6 % per survey [N = 109 surveys; 95% CI = 5-8% (calculated after Howell 1992)].
- A Kruskal-Wallis one-way analysis of variance (ANOVA; Howell 1992) indicated that the percentage of species that were listed but not quantified was different among years ($H = 10.0$; $p = 0.007$), but this difference does not tell much, as the mean percentage was always relatively low: 9 % in 1997, 5 % in 1998, and 6 % in 1999.
- In three reports, the number of taxa quantified was actually higher than the number of taxa observed, due to repetitions of taxa, with an average excess of 20% (N = 3 surveys; 95% CI = 5-34%). This could have been due to the same taxon being present on the surface, beneath, and/or within the substrate.
- Operationally, if a species is not listed as present in a specific habitat type in a specific year but has been documented as occurring there in previous years, it is assumed to not be present that year. There is no easy way around the problem of not knowing whether observers failed to see it and/or correctly identify it, or whether it really was absent from the site, as the potential number of species that might be present is very large and differs by habitat type. Zero abundance is therefore inferred from an absence of a positive abundance value at sites where a species has been previously shown to be present in that habitat type.

2.1.4 Recommendations

The errors listed above may affect the credibility and usefulness of some Shorekeepers' data. The following recommendations could improve future Shorekeepers' data collections:

(i) The main properties defining each habitat type have to be strongly emphasised in the Guide, in training sessions, and in instructing the volunteers at the times of the surveys. Improved emphasis might be done, for example, by using "bold fonts" for key terms in the text and/or by making a table that lists only key elements defining each habitat;

(ii) Even with the above improvements, some mislabelling of habitats might still occur, so there should be an on-going DFO database manager devoted to the Shorekeepers project during the survey season, who would review and correct such mistakes when field sheets are sent received;

(iii) Greater emphasis should be also given to the importance of (a) establishing enough relevant location benchmarks, (b) referring each baseline end point to at least 2 benchmarks, and (c) referring to the same distances between benchmarks and baseline end points on successive surveys;

(iv) Since the electronic versions of the maps are the ones that are going to be used regularly for data analyses, their quality needs to be maintained at a high level. Copies of the original sketch maps should not be reproduced at a smaller-than-original size, or the symbols, lines, and letters may become too small to read well. This problem occurred in the past when some of the sketch maps were drawn on two 8.5x14 inch sheets, necessitating some reduction in size prior to scanning;

(v) The importance of clearly noting the scales for both length and width in the map should be further stressed and checked by the database manager. Action to correct past, if possible, and potential future problems, as necessary, should then be taken;

(vi) To avoid potentially incorrect determinations of habitat areas, the database manager needs to determine the correct area for each habitat from the sketch maps provided by the survey teams, either electronically or otherwise;

(vii) Habitat slope and elevation can be only determined in the field; no technician will be able to correct mistakes once the data are produced. Thus, (a) it is essential that the determination of these two variables is clearly understood in training sessions. Also, (b) it may be useful to state in the guide the potential mistakes (see above) that might occur in elevation determinations, based on our analyses, to caution survey teams learn not to commit them; and

(viii) One of the main aims of the Shorekeepers project is to document the abundance of intertidal plants and animals precisely. The abundance of all taxa observed in a given habitat needs to be specified in the final computer-generated tables.

2.2 Degree of spatial consistency

2.2.1 Objective:

To estimate the consistency (and credibility) of measurements taken by survey teams.

2.2.2 Methods:

Comparisons were done for 4 sites between pairs of surveys separated in time by less than 2 months. These sites were (sampling dates in parentheses):

SK1997003, McGuffie Road, Nanaimo (26 May - 11 July 1999)
SK1997004, Entwhistle, Nanaimo (31 May - 12 July 1999)
SK1997009, Seadog Road, Nanoose Bay (17 May - 16 July 1999)
SK1997013, Driftwood Beach (14 July - 6 August 1999)

2.2.3 Results

2.2.3.1 Benchmarks

Problems occurred in two of the sites:

• *SK1997004*

Benchmark #4 was in the centre of a "rotten log" on 31 May (35 m to the baseline), but at the base of a "large fir tree" on 12 July (31.66 m to the baseline).

• *SK1997009*

Benchmark #3 was at 34.68 m from the baseline on 17 May, but at 33 m on 16 July. Benchmark #4 was at 24.9 m from the baseline on 17 May, but at 22.4 m on 16 July.

2.2.3.2 Habitat identification and form

Problems may have occurred at all four of the sites, although some perceived problems (it is impossible to tell from the data) may not be real if they are the result of subsequent algal growth, or disappearance,

during the time interval between samplings. Geological substrate classification differences are considered real discrepancies:

• *SK1997003*:
C1 (cobble): On 26 May, a 26-m^2 C1 band, perpendicular to the shoreline, was completely absent on 11 July.
R1-R2 (rock): On 26 May, R was a continuous area that was incorrectly separated into two areas.

• *SK1997004*:
C1: On 31 May, its outline was not mapped. On 12 July, there were two separate C patches, but only one was recorded.
R1: On 31 May, its outline was only partially mapped. On 12 July, there was only one R patch, but R1 and R2 were identified, instead.
OA1 (other algae): It was only present on 12 July, but incorrectly, since the percentage cover of *Fucus* then was reported to be 67%. It should have been named FA1 (*Fucus* algae).

• *SK1997009*:
R1: On 16 July, R1 included an area recorded as UA1 (*Ulva* algae) on 17 May. The creation of UA1, however, might be interpreted as incorrect, as there was only 31% cover of *Ulva* (a minimum of 50% is required to determine an UA habitat). However, >50% cover determined the habitat type, while an individual quadrate in that habitat may have less vegetative cover.
FA1: On 17 May, 51 m^2 were recorded as FA1, but as R1 on 16 July.

• *SK1997013*:
R1: On 14 July, R1 included what was identified as FA1 on 6 August. Additionally, on 6 August, R1 lost 138 m^2, which were then included in C1.
FA1: On 14 July, FA1 did not exist. The area identified as FA1 on 6 August was part of R1 on 14 July.
C1: On 14 July, C1 was only identified in half of the study area, but, on 6 August, C1 was a continuous area that extended from the left to the right borders of the sampled area. That represents a difference of about 288 m^2. However, in this case, algae covered half the cobble area.
OA1: On 14 July, OA1 was reported as covering about half the portion of C1 that was shown on 6 August.

2.2.3.3 Elevation

A posteriori data analysis calculated the variable *"% variation of elevation between dates 1 and 2"*, using the highest and lowest values for each habitat (except for the lowest elevations for habitats touching the water line at the time of sampling, which were not good estimators of a habitats' lowest limits, given that water line elevations varied among sampling dates). Results are shown in Table 3. Elevation values in the second survey could be between 5% lower and 26% higher (95% CI) than in the first survey. Differences may be real, depending on how the habitats were identified and the boundaries observed. 99 represents missing data, MIN and MAX are often reversed, so analyses (not the case here) like this should only be done with edited, corrected data.

2.2.3.4 Slope

The *"% variation of slope between dates 1 and 2"* was calculated (Table 3). Slope values in the second survey were between 90% lower and 31% higher (95% CI) than in the first one. Again, a value of 99 represents missing data, so analyses (not the case here) like this should only be done with edited, corrected data.

2.2.3.5 Habitat area

The *"% variation of area between dates 1 and 2"* was calculated (Table 3). Area values in the second survey could be between 161% lower and 428% higher (95% CI) than in the first one. Again, differences may be real, depending on how the habitats were identified and the boundaries observed.

2.2.3.6 Taxa identification consistency

For each habitat type from each site, the *"% of taxa identified in both dates"*, the *"% of taxa identified only in date 1"*, and the *"% of taxa identified only in date 2"* from the total number of taxa given for each habitat type in both surveys were determined. Results appear in Table 3. Of the total number of taxa identified in conjunction by both survey teams for each habitat type, only 39% (95% CI: 24-54%; N = 10) were given by both teams. This may be because rarer species did not occur in all quadrats, and hence might not appear to be in a habitat on different sampling dates. As discussed below, the majority of the variation observed was attributable to quadrat-quadrat differences, even when selected from the same area, year, habitat type, and habitat location.

The percentage agreement in species identifications between the two sampling dates changed when only easily-identifiable taxa with multi-year longevities were considered. The following 14 taxa, or taxa groupings were assessed: aggregate anemone, brown barnacle, common barnacle, coralline algae, encrusting red alga, *Fucus*, Pacific blue mussel complex, Pacific oyster, periwinkles, *Petrocelis*, purple star, *Sargassum*, shore crabs, and true limpets. Of the total number of these selected taxa that were identified for each habitat type by at least one team, only 59% (95% CI: 42-76%; N = 10) were given by both teams. Agreement between teams is thus slightly better when only easy-to-identify, long-lived taxa are considered, although the difference is not statistically significant between the two analyses (Confidence Intervals overlap) and the mean percentage of agreement is still moderate for the second analysis. A problem here is that different codes may be used for the same species. Coralline algae, pink crust, *Fucus*, and periwinkle have 2, 2, 3, 3 taxa codes depending on the different Latin names in the taxonomic list. i.e. periwinkle = 55 is *Littorina* sp. while others used 152 *Littorina sitkana* or 212 *Littorina scutulata*. The different categories need to be combined prior to analysis.

2.2.4 Discussion

Accurate habitat identification is one of the main requirements in the Shorekeepers database, but the reliability of survey teams to correctly identify habitats may not be consistent. Possible incorrect labelling of habitats (for example, naming OA1 as an area that actually has > 50% cover of *Fucus*) could be minimised by auditing for repeatability by survey crews in assessing percent cover. Potential subjective errors are impossible to detect afterwards, as individual quadrates are not a guide; by chance, they may be < or > 50% algae covered. Suggested inconsistent standards of mapping of habitats may require improved training of survey team leaders, as problems in benchmark positioning, habitat definition and in map drawing cannot be solved once field sheets are produced.

Another main objective of the Shorekeepers project is the listing and quantification of the benthic flora and fauna in each of the habitats. This is really the core process in measuring interannual variations of selected species or of the entire community, and in being able to relate such variability to a number of possible causes, such as human development, pollution, harvesting, or natural phenomena. However, survey teams showed differences in identification, likely in part because of inaccurate identification of species. These may also in part be the result of temporal variations in presence/absence of taxa between sampling dates, or the micro-spatial distributions of less-abundant species. Both latter factors would affect the likelihood of their being sampled, and hence the potential agreement between teams in documenting biodiversity between years. Biodiversity comparisons may thus be most meaningful only in the context of total numbers of species present that are readily identifiable.

The suggested low-to-moderate identification reliability compromises data analysis about potential community changes using 1997-1999 Shorekeepers data. Two ways to overcome species identification issues in the future are :

(i) to train people better in species identification, and/or

(ii) to identify only a selected list of taxa whose identification be relatively easy due to their distinct morphology, colour, size, or location.

The former will be difficult, as there are no resources for such training and in many cases, relevant taxonomic information may be unavailable. Regarding the latter, consideration of only a specified suite of species may fail to document important ecosystem changes that involve other species. In short, monitoring

ecosystem species diversity effectively requires that a capability to identify species exists, and that all species are subsequently monitored.

2.3 Accuracy of identifications of plants and animals

2.3.1 Objective:

To determine how accurately field teams identified organisms in their surveys.

2.3.2 Analysis:

At this stage, it is not possible to determine if identifications of intertidal plants and animals were done correctly, as no samples or pictures from the original surveys are available. One has to trust that the survey teams correctly used available field guides to the best of their capabilities, although it may not always be sufficient and equal. However, different results obtained by different teams surveying the same sites, even though there was about 2 months of difference between survey dates (see the previous section), suggests that survey teams may not be consistent in their identification of at least some organisms.

The level of taxonomic resolution for observed organisms was primarily limited by the instructions in the Shorekeepers' Guide, which advised, for example, that an organism could be identified to the species level, or the genus (*e.g. Stronglyocentrotus*), family (*e.g.* Cottidae), class (*e.g.* Turbellaria), or even phylum (*e.g.* Bacillariophyta) level, depending on the knowledge of the identifier. Survey teams generally followed this description system in their field sheets. The problem with such a system is that one team, for example, could identify a given sponge as "Porifera", but another team with greater training in marine taxonomy could identify the same organism by its correct species name (three species of sponge are currently listed in the Shorekeepers' Guide). Thus, any ecological comparison between years could be affected by the proficiency of the particular team that surveyed the site. This is a problem even within a habitat type in most surveys, as different individuals often compile data from different quadrates within the same habitat.

An additional indication of the quality of identifications was possibly given by the names that some survey teams assigned to organisms. For example, a particular team used names such as *Petrocelluas* or *Petrocellis* for *Petrocelis*, probous worm (for proboscis worm), red tuf turf (for red tufted algae), *Crasalaria* or *Grassalaria* (for *Gracilaria*), *Alva* (for *Ulva*), limpel (for limpet), *Rodomelia* (for *Rhodomela*), and *Entemorpha* (for *Enteromorpha*) throughout their field sheets, sometimes giving even different names for the same organism. This poor quality suggests that the survey team was not particularly familiar with the intertidal biota of their area. This lack of knowledge may have resulted in erroneous identifications. However, common and Latin names are usually enough to code the correct taxon number, regardless of spelling errors in the names.

2.3.3 Recommendations

The correct identification of every organism found in any community is a complex and difficult challenge. Even professional marine biologists usually find trouble in correctly identifying taxa that are not familiar to them, at least at taxonomic levels lower than phylum or class. Thus, it is normal to expect some higher level or erroneous identifications, particularly from people without experience in marine biology. However, since one of the main objectives of the Shorekeepers program is to document spatial and temporal patterns of variation in the abundance of intertidal organisms, it is important to ensure that identifications used in analyses are both accurate and meaningful. This may mean focusing only on species that are readily identifiable. For species identifications that Leaders are unsure of, they are now required to preserve representative specimens of the organisms. They are also tasked with confirming the report sheets are correctly completed, so if format errors arise, leader training may need to be reassessed.

Recommendations are:

(i) Support should be allocated to training knowledgeable leaders (preferably the same ones) across years to minimise data quality issues arising from new, relatively inexperienced leaders participating every year. Essentially, the better the quality of training, the better the results likely to be obtained.

(ii) Alternatively, if funding is limited or if survey leaders are renewed periodically, a more limited overall species identification of organisms is justified. This could be achieved by restricting identifications only to conspicuous taxa that are easily recognised because of their morphology or size. Depending on the taxonomic complexity of different groups, identifications should be specified to either species or to higher taxonomic levels, barring explicitly the possibility of multiple levels of identifications of a single species. This approach was in fact the one utilised in the first edition of the Guide in 1996, but it was dropped because different habitats and locations had such different biological community structures that to try and encompass them all, effectively met that a very large species identification list was required, one which would never functionally be complete.

Another approach might be to consider abundant species, keystone species, or biological complexes in detailed analysis, depending on the particular interest of the study. In preliminary analysis of the overall data, it is evident that even among common, easily recognisable species, there is much annual variability. Given this, the power of any statistical analysis of a species' abundance over time may not be very high. What may be more useful in most instances in the evaluation of the overall "environmental status" of a site is to have a suite of metrics (e.g. *about species abundance, species diversity, species spatial variability, habitat features, etc.)* that adequately describe this. What they might be is unknown at this time, but criteria used in the development of a freshwater Index of Biotic Integrity (IBI) (Karr and Chu1999) might be an example of what needs to be determined for the marine environment. Use of such a suite of appropriate metrics will require specific written descriptions and photographs/drawings for designated species. Less species to identify might also leave more time for replicate quadrat samplings during surveys.

2.4 Transcription error rate between field sheets and computer spreadsheets

2.4.1 Objective:

To determine the error rate in the transcription of information from field sheets to computer spreadsheets.

2.4.2 Methods:

The information that was recorded in field sheets by survey teams was compared with the information that finally appeared in computer spreadsheets. The ideal transcription of data from field sheets to spreadsheets should involve no errors. However, this is an unrealistic assumption, even for professional biologists, so the transcription error rate was estimated for a random sample of surveys.

Possible differences in the information that needed to be transcribed (*e.g.* dates, site location, habitat description, etc.) were determined by visually comparing values between field sheets and computer spreadsheets. However, to determine possible differences in the information that needed to be treated mathematically before entering it in spreadsheets, the necessary calculations were repeated using data from the field sheets and compared with the data that appeared in the spreadsheets. This was done for the mean abundance of animals and plants, for which averages for a sample of taxa were recalculated. For each taxon, averages should be obtained from the abundance recorded for the different replicate quadrats that are normally surveyed for each habitat type. When there was only one patch of a certain habitat type at a given site, determining the average abundance for each taxa was simply done by calculating the arithmetic mean from the replicate quadrats. When there were two patches per habitat type per site, however, a weighted average for each habitat type was derived, taking into account the different patch sizes.

For example, for a sessile organism (for which abundance is expressed as percent cover) occurring in a habitat that consisted of two patches, mean abundance (C = mean percent cover) was estimated as:

$$C = [(C_1 A_1) + (C_2 A_2)] (A_1 + A_2)^{-1} ,$$

where:
C_1 = mean percent cover in patch #1 (based on all quadrats surveyed in patch #1),
C_2 = mean percent cover in patch #2 (based on all quadrats surveyed in patch #2),
A_1 = area of patch #1 , and
A_2 = area of patch #2.

For a mobile organism (for which abundance is expressed as density) occurring in a habitat that consisted of two patches, mean abundance (D = mean density) was estimated as:

$$D = [(D_1 A_1) + (D_2 A_2)] (A_1 + A_2)^{-1} ,$$

where:
D_1 = mean density in patch #1 (based on all quadrats surveyed in patch #1),
D_2 = mean density in patch #2 (based on all quadrats surveyed in patch #2),
A_1 = area of patch #1 , and
A_2 = area of patch #2.

For each patch, mean density was calculated as the mean number of organisms counted (using data for all quadrats surveyed in that patch) divided by quadrat area (all area calculations were done in m^2).

Mean abundance for selected taxa of the 135 in total recorded were calculated from a sample of sites (N = 15) that were selected at random for each year. The selected taxa (*Fucus*, *Ulva*, *Littorina*, *Hemigrapsus*, and *Pisaster*, to name a few) are common in south-eastern Vancouver Island, so this analysis is representative of the most conspicuous intertidal organisms. Due to their importance, the abundance of these organisms was calculated for more than one site.

Results: In general, the transcription error rate of information between field sheets and computer spreadsheets was low when the information just needed to be transcribed. For example, dates of field surveys were incorrectly entered in spreadsheets in 4 % of the cases (1 in 1998 and 3 in 1999; N = 109), habitat areas were incorrectly entered in 3 % of the cases (2 in 1998 and 1 in 1999), and baseline length was incorrectly entered in 2 % of the cases (1 in 1997 and 1 in 1999). With a designated DFO person to maintain the database, these errors would be expected to be identified and corrected.

Errors were more common when the information needed to be treated mathematically before entering it into spreadsheets (such as with the mean abundance of organisms). The first type of transcription error relating to abundance was that some of the taxa originally cited as present were not all incorporated into the tables describing their abundance.

The second type of error was that the average abundance of taxa were sometimes incorrectly recorded. There was a relatively large difference between calculations done for one-patch habitats and those for two-patch habitats. For the former, the transcription error rate was 25% (95% CI: 19-32%). For the latter, however, the transcription error rate was significantly higher at 82% (95% CI: 72-92%). There were little differences between years, indicated by the small confidence intervals for both cases.

2.4.3 Recommendations

2.4.3.1 Maps:

The scanning technique needs improvement, since maps are essential to assess interannual differences in habitat characteristics. Digitised maps would allow between-year comparisons and accurate habitat area measurements.

2.4.3.2 Data transcription:

The high transcription error rates detected for abundance data indicated that the current method followed for data input (*e.g.* using a variety of people with different levels of training) has to be improved. To

eliminate or minimise transcription errors in the future, a solution might be to assign a designated full-time DFO data technician to input these data and to manipulate them appropriately.

It should be noted that calculated abundance data generated by field surveys between 1997 and 1999 may need to be recalculated after the errors in the original data were corrected in 2001, so they can be used with mathematical confidence. A standardised worksheet should be developed to help shorekeepers in the computation of abundance when multiple habitats are involved.

2.5 Field audits

2.5.1 Objective:

To determine the reliability of survey teams to identify habitats and to estimate habitat area, elevation, and slope.

2.5.2 Methods:

It was initially hoped to undertake the following: "an audit survey of approximately 10 of the sampled sites on the east coast of Vancouver Island will be conducted in August 2001. For this part of the work, an NGO group will perform a survey and, a few days afterwards, the site will be resurveyed by more experienced and fully trained people (e.g.: biology coop students)." However, for a variety of logistic reasons, it was ultimately decided that the audits would be led by trained biologists already involved with the Shorekeepers program. Even this effort encountered logistic problems, though, and the data ultimately available was a series of six audits (in six different sites) during August and September, led by two different persons. Time gaps between the original survey and the audit were approximately 1.5-2 months, for four sites, and 0-2 days for the other two sites. Sites and dates are given in Table 4.

2.5.3 Results and Discussion

2.5.3.1 Time gap and habitat modification:

The Shorekeepers' Guide defines habitats by a combination of substrate composition and the abundance (% cover) of frequent organisms such as rockweed, eelgrass, sea lettuce, and other macroalgae. Thus, a given intertidal patch might be identified differently depending on the season, as the above organisms vary in biomass abundance seasonally. Therefore, an adequate audit program should be based on time gaps between the original survey and the audit of not more than a few days. With time gaps of 1.5-2 months, it is possible, and even likely in this assessment, that differences in habitat identification would result not from an incorrect identification by the first survey team (which is what an audit program aims to detect), but simply from a natural change because of the seasonal growth of some relevant organisms.

The percentage of habitat patches that were only identified in one of the two surveys at each site with a long time period between samplings, by site, was 0% (Willows Beach), 20% (Cadboro Bay), 40% (Moses Point #1), and 78% (Ten Ten #2). For each site, temporal differences were always dominated by some kind of vegetation in either surveys. In no case were temporal differences due to changes from one type of "physical" habitat (e.g. rock) to another (e.g. sand). This strongly suggests that temporal differences in habitats identified resulted from the seasonal growth of vegetation cover rather than from identification errors by the first survey team.

On the other hand, the absolute abundance of percent cover of plants that define "biological" habitats (*e.g.* rockweed, eelgrass) that might have helped to compare temporal differences in habitat identification are unknown for audit dates. The abundance of these plants is simply noted as greater or lesser than 50% cover, as indicated by the habitat code used. It was hoped that photographs of the quadrates could be taken, but this was not logistically feasible.

An additional problem of these temporal differences in patch names was that typically, "biological" habitats noted in a given survey may have occurred in an area that was part of two or more other different habitat types in the other survey. This prevents us from undertaking a useful comparison of relative area, elevation, and slope for habitats that occurred in the two surveys where only some fraction of their areas were

common between the two surveys. At this stage, it is impossible to know whether the maximum or minimum elevation of, for example, S1 (sand habitat #1) occurred in an area that was identified as S1 in both surveys or not. Possible elevation differences between surveys could be attributed to an incorrect determination by one of the teams, or differences could be real and be for different actual intertidal areas. A similar problem applies to area comparisons: we cannot make valid comparisons between surveys for a given habitat type because areas will often differ due to the existence of "biological" habitats somewhere in the surveys.

2.5.3.2 Habitats useful for comparisons:

Among all the habitats identified in the 12 surveys/audits considered here, only seven were unaffected by seasonal temporal changes in "biological" habitats (i.e., these habitats looked similar between the different surveys maps) and were not in contact with the waterline at the time of at least one survey (this would prevent their useful comparisons between surveys, as the waterline varies between surveys, affecting area estimations). This limits the representativeness of this audit program. Nevertheless, comparisons between surveys, using those seven habitats, are as follows:

Habitat area: Differences between the area estimated by the initial survey team and the later audit teams were variable. On average, audit teams reported values 67% higher than survey teams (95% CI: -63% and 197%; N = 6 -one area value was not given-).

Habitat elevation: Differences between habitat height as estimated by survey and audit teams were also quite variable. All reported height values were standardised to the zero level of the tide table (chart datum) to make valid comparisons between dates. On average, audit teams reported values 20% higher than survey teams (95% CI: -39% and 78%; N = 7) for the minimum height of each habitat and **25%** higher (95% CI: -9% and 60%; N = 7) for the maximum height of each habitat.

Habitat slope: None of the audit teams measured habitat slope due to a recent change to the protocol. However, habitat slope is still a required number to allow data entry.

Differences between audit teams: One of the necessary conditions of an accurate audit program is that audits be done by the same team or that they be led at least by the same person. If done by two different teams or led by two different persons, an additional source of variation is introduced (Krebs 1999). If this problem exists, it can not be determined if observed data differences between sampling dates for a given site resulted from errors committed by presumably the first survey team, or simply from proficiency differences between the two audit teams. For this audit program, the assumption that both audit teams worked exactly in the same way is unwarranted, as no formal "agreement" test has been done to prove it. In addition to what relevant textbooks say about this matter, common sense indicates that a certain degree of difference will exist among teams, particularly when they are formed by several people.

In summary, under the audit program that occurred in this study, differences between data obtained on the two sampling dates for each site could have resulted from:

(i) errors committed by the first survey team,

(ii) seasonal differences in the abundance of intertidal vegetation, and/or

(iii) proficiency differences between audit teams.

2.5.3.3 Recommendations for Future Audits

The evidence discussed in this section, particularly based on the seven habitat surveys referred to above, and from data discussed in section #2, indicated that mistakes are occurring in the determination of habitat names, areas, elevations, and slopes. It is unclear whether the frequency of mistakes has decreased in recent

years because of improved training of leaders. Recommendations for future audits to overcome the problems shown by this evaluation are:

(a) audit teams should consist of well trained staff;

(b) audit about ten sites from relatively distant areas, so the audit program would be representative of different survey teams that are operating on Vancouver Island;

(c) for each audited site, do the audit within a few days after the original survey, preferably during the same series of low tides;

(d) have only one team (with only one leader) do all of the audits of other teams; and

(e) objectively quantify the abundance of organisms that define "biological" habitats, as the identification of those habitats directly depends on knowing their abundance. *These data will be most* relevant when percentage cover is near 50%, as this is the decision point for determining whether a habitat might be, for example, a rock or *Fucus* habitat, depending on the amount of *Fucus* present. Digital photographs (minimum of six; but more if time and electronic storage space are available) of randomly selected locations where percentage cover could be calculated may be an appropriate approach.

(f) A standardised worksheet should be developed to help shorekeepers in the computation of abundance when multiple habitats are involved

3 Survey Design and Data Analysis Evaluation

3.1 Design Issues

The Shorekeepers' Guide survey approach is an opportunistic survey design where areas to be surveyed are selected based on the desire of non-professionals to obtain standardised, credible data to document and to evaluate changes over time. One goal of the project is to obtain much more data and over a wider scale that could be gathered given the constraints of resource management agencies.

At each survey location, permanent study areas are delineated. Within each study area, habitat units of at least 25 m^2 are identified. Within each habitat area, transects that cross the habitat perpendicular to the shore are located and quadrats selected on each transect line. Information is collected for each quadrat on species abundance.

3.1.1.1 Opportunistic sampling of sites

Because of the opportunistic selection of survey sites to date, it is difficult to know how to extrapolate the results of these surveys to a larger geographical area, or ecological unit. Among all sampled sites, there are grouping that can be made that have similar "environments" (fetch, summer thermal stratification, etc.) and that fairly represent a given "larger" area or regional context, but these are all largely in relatively unimpacted areas. There is thus nothing to compare them to in the short-term, although in the long-term, they represent a baseline with which to compare potential future changes against. Sites surveyed may have been selected because they are easy to access, they are close to population centres, or they were near a potential environmental impact (e.g. refer to Step 2 of the Guide). Trends seen in these opportunistically selected sites may also not be indicative of trends in the larger ecological unit because of potential confounding factors (these sites were not randomly selected) present at these sites may not be present elsewhere.

If a site has been chosen because of community concern about a potential environmental impact (e.g. the construction of a new thermal generating station), the community groups should be aware of the need for a minimal BACI (before-after-control-impact) design so that changes over time not related to the impact can

be distinguished from changes over time related to the impact. There have been several different BACI designs proposed (e.g. Underwood, 1991) and it would be unfortunate for a community group to collect data that cannot be used to meet their own concerns. The Shorekeepers' Guide (Step 2, paragraph 1) alludes to the need for such careful design for impact studies. To date, no specific study has been undertaken with Shorekeepers. Probing as to intent and expectations needs to be done by the DFO representative when contacted by a community group desiring to undertake a specific study. The purpose of each survey should be recorded in the database; collection of some of this information has been planned for on the first page of Form 1.

On Form 1, space is available for the Registration Number of a nearby potential "Control or Treatment" study area. This is confusing as there may be more than one "Control" or "Treatment" site, and the word "Treatment" implies some active, planned, intervention at the study area. The term "Impacted" site may be clearer as the listing of potential objectives for the study (Figure 3 in the Guide) includes both planned and unplanned interventions. The sample data file evaluated did not include this information.

The field "Reasons for Choosing this Study Area" allows a single selection from the list in Figure 3. One choice is "In or Near Human Impacted Area". Note that here the terminology is now "impacted area" rather than "treatment area", but the use of "human impacted" may be confusing for some choices for the objectives. For example, a study group may wish to study the differences near and far from a freshwater stream. It is unclear how the objectives and reasons would be chosen for such a study.

Step 13 describes the delineation of habitats within the study area. Each habitat is to be a minimum of 25 m^2 and if there are multiple habitats, each is given a unique number. However, there are no explicit instructions if all instances of habitat units are to be sampled in each survey area, i.e., if there are two S habitats, S1 and S2, are both to be surveyed? This is the intention, though. If time does not permit measuring both habitats, how is this recorded in the database? Which habitat units should be selected if all instances cannot be surveyed? If some habitat units disappear over time (refer to the example at the end of Step 13), how is it recorded in the database that a particular habitat unit no longer exists, or if it simply wasn't sampled because of time constraints?

3.1.1.2 Habitat Determination

Another key issue for the Shorekeepers Project is that naming some habitats by biological cover (e.g. UA habitat, containing > 50 % cover of *Ulva*) may cause problems because the defining species (e.g. *Ulva*, for UA) often changes in abundance over years, but many other species in the same habitat may not. Thus, for example, a spatial part of the same UA unit in year 1 may be included in a different habitat unit in year 2. The new habitat names might suggest that big changes are occurring in the intertidal, but perhaps only the defining species is changing significantly, which could well be the case with *Ulva*. A solution could be to always analyse the same permanent patches for a given site, regardless of which species predominates at sampling times, but this has problems with some habitat types because loose substrate characteristics often change over time and may move round spatially.

The restriction to habitat units of 25 m^2 or larger implies that no information may be collected from survey areas containing aggregates of small, patchy, habitats where no one substrate predominates. These areas may be of considerable ecological interest as much can happen on the margins of the habitat. However, there may be no way to modify the protocol to include these areas without introducing considerable problems in the data collection and analysis.

3.1.2 Permanent monitoring sites

The survey areas are located using a permanent baseline. As such, the same study area may be surveyed over several years.

Green (1979) discusses the advantages and disadvantages of permanent monitoring versus temporary monitoring sites. The primary advantage of a permanent monitoring site is the elimination of one source of

variation when searching for changes over time. This is particularly true when the site-to-site variation is large relative to other sources of variation.

A method to investigate the "causes" of the variation in the data is a variance component decomposition. A variance decompositon partitions the total variation observed in the data to various sources. As a very simple example, suppose there were two areas with the following counts for 5 quadrats at each area:

Area 1	Area 2
5 6 7 4 9	15 17 13 16 14

The "Total Variation" is defined as the variance of the 10 numbers all combined together ignoring the area delineation, and the maximum likelihood estimate is 21.84 for this example. This simple example has two sources of variation - random quadrat-to-quadrat variation within each area, and area-area variation. A variance component procedure would partition the total variation into:

Variance Component	Estimate
Var(area)	18.74
Var(Error)	3.10 , i.e. quadrat-to-quadrat error.

Notice that the individual variance components add to the total variation, i.e. , 18.74+3.10=21.84.

The presentation of the decomposition can be improved by expressing the decomposition in percentage terms so the decomposition is independent of unit changes (i.e., imperial/metric units). The percentage variance decomposition is:

Var(area)	18.4/21.4	=	86%
Var(quadrats)	3.1/21.4	=	14%

The decomposition in the above example shows that most of the variation in the data arises from area-to-area differences rather than among quadrats within an area. The variance decomposition can be extended in more complicated surveys to multiple levels, i.e. areas, year within area, habitat types within years within areas etc.

A variance decomposition analysis of the density values, averaged over all species was performed using the provided data and the results are summarised in Table 5. Surprisingly, Table 5 indicates that the majority of the variation is attributable to quadrat-quadrat differences even when selected from the same area, year, habitat type, and habitat location. However, the minimum and maximum values of the decomposition indicate that while most species follow this general pattern, they can vary considerably in how their variation is partitioned. As an analogy, the earlier example values would need to take the following pattern to have a similar large among quadrat variation as opposed to an among area variation:

Area 1	Area 2
5 15 7 13 9	6 17 4 16 14

The implication of the high variation attributable to quadrats is that the gain from permanent monitoring stations is not expected to be great! Because there does not appears to be large variation among the different areas, this indicates that the opportunistic sampling is choosing similar sites. And there appear to be only small changes over time (as measured in the data collected). The relatively small variation over replicate habitat types in the same area indicates that perhaps the protocol could be modified to relax the requirement that multiple occurrences of a habitat be measured and that more quadrats should be taken at each habitat type.

Surprisingly, there was little additional variation caused by habitat type. This may be an artefact of the species and habitat types present in the database evaluated.

All of this implies that with the large "noise" in the data, it may be difficult to detect changes over short time periods and extended data collection over a number of years should be planned for.

3.1.3 Laying out transects and quadrats:

According to the Shorekeepers' Guide (Module 1, Step 27), transect lines are to be placed "across a habitat unit" and the procedure for deciding upon the locations of the quadrats ensures that these are uniformly located and away from habitat edges.

There is some ambiguity in the description of the placement of the transects - the diagrams always show them perpendicular to the water, but this doesn't appear to be explicitly mentioned other than in Step 27 that the "transect lines will run the width of the habitat unit". The term "width" is defined earlier (Step 12) to mean parallel to the water edge, but an explicit mention at this part of the guide may be useful.

The quadrats are located by dividing each transect line into equal portions and selecting the interior points.

The net effect of this procedure is that the quadrats will tend to sample preferentially the interior of a habitat unit. If the organisms are located uniformly over the habitat unit, this will make little difference; however if organisms tend to be differentially located at the borders of habitat units, they will tend to be missed by the sampling protocol.

The instructions require that the same number of quadrats be sampled on each transect. This will result in a uniform distribution over the interior of the habitat area and will help avoid sampling preferentially in areas of higher density. The instructions for choosing the number of quadrats per habitat area allow effort allocation to vary among habitat types and users may be tempted to apply the same within the habitat area. The rational of why different sampling strategies are used within habitat areas versus among habitat areas needs clarification in the Guide.

Concerning how many quadrats should be sampled, the guide indicates that Shorekeepers should sample at least 6 and not more than 15 quadrats per habitat unit.

More guidance could perhaps be given in choosing among the various combinations of number of transects and quadrats per transect. For example, is 2 transects x 3 quadrats per transect favoured over a 3 x 2 allocation? How should a shorekeeper choose between a 2x6, a 3x4, a 4x3, or a 6x2 design for 12 quadrats? However, it will likely make little difference given the large variation among quadrats noted above.

There is some flexibility given to samplers in the placement of the quadrats, i.e. within 30 cm of each point. Again, it should be emphasised in the documentation that the choice should not be dictated by "trying to get lots of data".

The manual states "If you want the results to be comparable across years, you need to follow the same procedure on sampling over time. This means sampling the same number of quadrats in all habitats of the same type in all survey years". While consistency is nice, there is no statistical requirement for uniform sampling over time. Indeed, based on a long time series, it may become desirable that certain habitat types should have increased sampling effort over time, particularly if the variability observed is unacceptably high because of limited samplings.

Some instructions on what to do with problems should be more clearly identified. For example, refer to Step 31 of Module 1. The instructions for avoiding double counting are "buried" under the "What if there are hundreds of individuals" heading.

3.1.4 Impact of repeated sampling over time

A community group may wish to repeat the same sampling protocol over time. The use of permanent monitoring sites was discussed earlier. However, given the tendency to use the same sampling effort and

the instructions on the placement of transect lines and quadrats, the same general locations within a habitat area may be revisited over time. Because of disturbances at the quadrat measured and around this quadrat, this may induce a human disturbance into the measurements, e.g. the impacting of the same general area over and over again.

Introduction of some randomisation might avoid this issue, but it would seem unlikely to be a serious problem overall. Quadrats sampled are relatively small, and given that habitat boundaries are somewhat subjective, it would seem unlikely that quadrats on different surveys would overlap much. Also, sampling is non-destructive, and even when soft substrates are dug up, this is a fairly dynamic habitat with much natural disturbance from wave action. Samplings are at the most repeated annually, which is assumed to leave sufficient time for a 25-by-25 cm area to recover biologically.

This overall issue was one that the Guide's authors gave considerable thought too, as there was debate among them as to whether to sample abundances of species in quadrats or simply to emphasis habitat mapping, possibly over a larger area than the 5 to 100 m area shore length recommended in the present Guide's protocol. In some respects it is a philosophical debate, and one that is dependent on how data obtained might be used by resource managers. This is likely to always be an issue until a specific problem is identified, at which time the merits of the contrasting approaches can be specifically applied.

3.2 Data collection and analysis

Some sample data from the database was provided in the form of an Excel spreadsheet. As noted earlier, information on the objectives of the study and why a particular study area was chosen were not present on the spreadsheet for a particular area.

3.2.1 Large counts:

A review of the dataset provided indicated that there are a few records with extremely large counts, e.g. a count of 1,000 for species 44 (hermit crab); counts of over 1,000 and as high as 3,000 for species 55 (periwinkle). It seems unlikely that these are actual counts and are likely estimates as noted in the guide. This could make analysis problematic as a few large counts have an extreme influence on means and standard deviations. Rather than trying to get an accurate count for particularly abundant species, they should perhaps just be coded as "very abundant". The proportion of quadrats with "very abundant" species in the survey could then be monitored.

3.2.2 Depth of excavation

The instructions in the Guide (Step 29) indicate that organism are to be counted in three substrate locations (S, B and W). The counts on the surface or beneath moveable objects are straight forward, but the counts of objects within the substrate may be problematic. The manual does mention that all quadrats should be dug to the same depth. However, this information does not appear on the data file and it is not clear at all how to "standardise" any of the data after the fact if digging was to a different depth, even if the depth and counts were known. Use of a single excavation depth consistently can avoid these problems.

3.2.3 Derived variables

It is difficult to determine if derived variables (e.g. density is a function of count and size of quadrat) are stored in the database or are recomputed as needed. It is recommended that derived variables not be stored in the database as it is too difficult to enforce consistency if changes and updates are made to the base values, e.g. how is density updated if a count value is changed?

3.2.4 Missing data/zeroes

A serious problem is the treatment of missing values and zeroes. This occurs at several levels:

- the habitat unit level,
- the quadrat level, and
- the species within quadrat level.

As noted in Step 13 of the Guide, some habitat units may disappear over time. Some habitat units may not be surveyed because of time constraints on the field crew. How are these denoted in the database?

The database has a field for the number of quadrats nominally surveyed. However, as seen in Table 6, the number of quadrats actually present for a survey does not always match the number nominally surveyed.

In some cases, the number of quadrats surveyed exceeds that nominally done. This may indicate that the nominal survey number is incorrect. More serious is when the number of quadrats present in the database are less than the number nominally surveyed. The database does not provide any information as to why a quadrat did not appear in the database. Whether it was because it was not surveyed (e.g. lack of time) or was it because no organisms of any species were observed is not obvious.

At the species level, the database does not appear to have a mechanism for entering a count of zero for a species in order to distinguish between a value of missing, because not counted, or not observed. For example, if there is no record for a periwinkle in a quadrat, is this because there was a count of zero or because the shorekeeper group did not see the periwinkles. This is most likely to be an issue for smaller and/or cryptic organisms. As an illustration of the scope of the problem, Table 7 shows the distribution of the number of quadrats with periwinkles recorded compared to the actual number of quadrats surveyed given that at least one quadrat recorded a periwinkle. Even for this common species, there are many quadrats that have zero counts but are not recorded in the database.

As a note in passing, Table 7 also indicates that there are some surveys where the nominal quadrats measured is six, but seven or eight quadrats are present in the database. It is recommended that with data extraction from the database where zeros hadn't previously been included, ie. for data entry prior to 2002, that a zero be put in for a count when the quadrat record is missing for a species.

Both types of missing records (quadrats and species) have implications when computing statistics on these species. Most statistical packages will not automatically impute a value of zero when a record is missing. Consequently, the mean density computed may be based on non-zero counts only. In an extreme case, the average non-zero density could remain constant, but the overall density could be in decline.

The Shorekeepers' Guide perhaps should be revised to include a checklist for the common species in a region. It is not clear if such a list is needed for rare species. On one hand, more of the counts will be zero, but on the other hand, the shorekeeper may not be able to recognise a rare species and DFO may not want these rare species to be removed for later identification, as suggested in Step 30.

The database should also be redesigned so that these zero counts are explicitly included when data are extracted from the system. Either the zeros should be included with the raw data (increases storage space, but no subsequent processing required) or the extraction programs should insert zeros when extracts are selected.

This has obvious consequences when estimating abundance of a species - the zero counts need to be included. In Module 2 (Information Management), Appendix B and C describe how abundance is computed for non-motile and motile species. In Appendix B, it is explicitly stated that averages are computed "including those quadrats where no observations were found". However, Appendix C is silent on including quadrats with zero counts.

3.3 Analyses

3.3.1 Changes in mean abundance

Presumably, the main purpose of these analyses is to detect changes in abundance over time rather than changes between areas or habitat types etc.

The usual analysis to detect changes in the mean is ANOVA. Even ignoring the opportunistic nature of the sampling design, analysis of these data presents many challenges:

- *missing cell combinations*. Not every area is measured every year; nor may each habitat type be present in each area etc. These types of designs are extremely difficult to analyse (Milliken and Johnson, 1984, Chapters 13-15) but modern packages such as SAS V.8 (Proc Mixed) can provide a comprehensive analysis. Excel and other simple packages are not up to the task;
- *different goals for selection of study areas*. For example, some study areas may be deliberately chosen as part of a BACI environmental impact study (see for example, the sample Form 1 in Step 8 of the Guide), while others may be a survey of specific habitat types. Before extracting the data for an analysis, the reasons for choice of selection of sites should be known. These will have a direct bearing upon the focus of the analysis. For example, in a BACI environmental impact study, the interaction effects between time and class of study area (control versus impact) are evidence of an environmental impact. In a "survey", the time trends, averaged over areas, would be of most interest;
- *weighted or unweighted analyses*. If interest lies in a BACI comparison for a particular habitat unit, it is unnecessary to weight the quadrats measured using the habitat area. However, if a large scale trend analysis is of interest, a weighting using habitat area will be necessary;
- *effective graphics*. Bar charts and similar graphs will likely not be sufficient to present a simple summary of the trends. This suggests that plots showing trends in abundance over time may be more suitable, perhaps separated by habitat type;
- *non-normality of response*. Counts are often modelled by Poisson random variables. However, there is likely to be overdispersion present in the data, i.e., many species are aggregated rather than spread uniformly. Such aggregation would result in the large variance component attributable to quadrats seen earlier. Variation in response also is likely a function of the mean. A transform (such a log or square root) may assist in the analysis. Alternatively, modern software such as Proc Mixed of SAS allows variances of observations to be function of means; and
- *many zeroes*. As noted earlier, there are many quadrats with zero count. Many organisms are present at low densities and are highly aggregated which gives many zeros and a few large counts. In such instances, large sample sizes will be needed to detect changes over time.

Fortunately, it appears that the relatively large variation attributable to variation among quadrats in the same habitat unit in a survey area will imply that consideration of autocorrelation over time or space may be moot and in many cases can be "ignored" without greatly compromising the results.

A sample unweighted trend analysis along with summary plots using the periwinkle data (after performing an 'ad hoc' insertion of the zero counts) is given in Appendices 1 and 2. The estimates showed a reduction in mean abundance over time by about a factor of 3 to 4, i.e., the mean abundance in 2000 was about 0.3 to 0.25 that in 1997 [Note: these results should be treated as preliminary as the objectives for each survey are ignored and the problem with zero counts should be properly corrected. *Littorina* sp., *L. sitkana*. and *L. scutulata* should also be combined prior to analysis.] The information present in such an analysis could be used for a power analysis to determine how many additional years or study areas would be required to detect a difference, but was not done here because of the *ad hoc* correction for zeroes.

3.3.2 Species composition

The above analyses could be repeated on a species-by-species case, but changes in species composition are also of interest. In this case, data will have to be converted from raw counts to relative proportions. This can be done either on a organism basis or some "weights" need to be developed to convert on a total biomass basis or on an "importance" basis.

The treatment of missing values/zeroes noted above will be important for a compositional analysis. In particular, a species that was not searched for should be treated quite differently than if a species had a count of zero.

Such analyses should be confined to a particular habitat unit, i.e. comparisons of species compositions among habitat units would not be informative and may simply reflect habitat effects.

Another types of analysis for species composition would be an analysis of the ratio of two species within quadrats (e.g. types of crabs). This could be handled (after a suitable transform) using ANOVA methods as described earlier.

Several methods of analysis are described in Digby and Kempton (1987).

3.4 Recommendations from a data analysis perspective

1) It is recognised that for areas surveyed to date, the focus was on method development and evaluation rather an assessment of "ocean health" *per se*. Sites were chosen more for ease of access rather than their utility in addressing specific ocean management issues. Future site selections should now focus on the latter, with clear identification of both "control" and "impacted" areas and appropriate replication of both.
2) With the documented large variance in the quadrat data, it may be difficult to detect changes over short time periods and extended data collection should be planned for. How long a site should be sampled is situation dependent.
3) There is no statistical requirement for uniform sampling over time, and so assuming that a long time series will be established, it may become desirable that certain habitat types (sensitive ones, those of particular interest, etc.) should have increased sampling effort over others over time.
4) The existing database should be redesigned so that zero counts are explicitly included when data are extracted from the system. Either the zeros should be included with the raw data (increases storage space, but no subsequent processing required) or the extraction programs should insert zeros when extracts are selected. Future data entries should include zeros.
5) Following collection of data at appropriately selected sites, a power analysis to determine how many additional years or study areas would be required to detect a difference should be implemented. This was not done here because of the *ad hoc* correction for zeroes, and because sites surveyed were not selected to address specific questions. However, it must be remembered that this is not an academic project. We simply don't know when, or even if, significant differences, or changes, will occur, so if we want to use the Shorekeepers program to monitor possible changes in intertidal ecosystems, we may need to monitor sites almost yearly, regardless of any 'academic' power analysis.

4 Discussion

The Shorekeepers Program has the potential to generate considerable data from volunteer groups. In order for these data to be usable, standardised survey protocols have been established. Efforts are typically made in the development of survey protocols to anticipate potential human errors and logistic survey and analytical difficulties that might arise. However, following collection of data over a few sampling periods, it is often useful to critically review how a program is working. The two reports that analysed the Shorekeepers' Guide survey and data reporting procedures, and the data analysis associated with it, respectively, which provide the bulk of this manuscript, both identify problem areas and provide constructive comments as to how they may be rectified. It is not always possible to correct past survey data, but in some instances this may be possible. The DFO authors of this manuscript have themselves in recent years identified and addressed some of the problems pointed out in this comprehensive analysis of the early data collection period, but there are others that still need to be addressed. More insertions of statements saying "caution – check data accuracy thoroughly" at appropriate locations within the Guide would particularly seem appropriate, or alternatively, request easier tasks for completion by shorekeepers.

It should perhaps be pointed out that the problems identified in this manuscript are primarily focused around the maintenance of data quality and improvements in data analysis. They do not address the overall utility of Shorekeepers' data for use by either resource managers or as a scientific database, i.e. are manager's likely to react and restrict actions if the abundances of non-commercial species that have no

obvious major ecological role are varying because of human impacts on the environment. Data utility will significantly influence the locations sampled, the data time series, and the further determination of what are the most relevant metrics to evaluate the "health" of regional intertidal ecosystems. Preliminary data are needed to be able to analyse how data might best be used, and what data specifically should be collected. This latter analysis has begun, but it will take a number of years and the incorporation of other information (e.g., the biological tolerances of the different species being observed to different pollutants, what guilds of species are most representative of "pristine" conditions, etc.) to complete.

The major issues that need to be re-evaluated in Shorekeepers are:

1) For any planned formal analysis, more information as to why a survey area was selected needs to be included. The analyses for a BACI design is quite different than an analysis to find a trend over time based on sites selected for general survey work;

2) The merit of changing the focus from organism-based habitats (e.g. UA, FA, OA) to permanently marked habitats, regardless of which species occur there; and

3) The merit of reducing the number of species sampled , focusing on those that are easy to identify, so that shorekeepers will make less identification mistakes and will have more time to take more replicates. This might involve the preparation of a revised field guide, with clear pictures of species in their different life stages and forms.

5 Recommendations

1. That Shorekeepers' data be verified by DFO, with corrections and/or modifications, as necessary, made before entry into the main database. A designated DFO Science technician should be tasked with this responsibility as the Shorekeeper database manager.

2. That the Shorekeepers' Guide be modified as necessary to address issues raised here.

3. That resource managers assess the appropriateness of the species array being sampled for their decision-making requirements. If it is not presently appropriate, then changes in survey protocols or analyses be carried out. Specific species, guilds of species, etc. may only be particularly relevant, and it may be shown to be desirable from a survey perspective to concentrate less on "rare" species and more on acquiring replicate samples for these species/complexes. The debate between concentrating more on habitats and less on species' abundance should be revisited in this review.

4. That audits of performance be a routine part of the leader training program, with a specific reference site(s) designated for survey by different teams within a relatively narrow time window (e.g. monthly tide cycle).

6 Acknowledgements

Financial support for the data reviews conducted by Drs. Scrosati and Scharwz was provided by the Environmental Sciences Strategic Research Fund. Funding for the original development of the Shorekeepers Guide was provided from the Oceans Act Implementation Fund.

7 References

Digby, P.G.N. and Kempton, R.A. (1987). Multivariate Analyses of Ecological Communities. Chapman and Hall: London.

Green, R.H. (1979). Sampling design and statistical methods for environmental biologists. New York: Wiley.

Howell, D.C. 1992. *Statistical methods for psychology*. Duxbury Press, Belmont, 693 p.

Jamieson, G.S., C.D. Levings, B.C. Mason, and B.D. Smiley 1999. *Shorekeepers' guide for monitoring intertidal habitats in Canada's Pacific waters*. Fisheries and Oceans Canada, Pacific region. Modules 1, 2, and 3. 1 Vol. (Looseleaf). http://www.pac.dfo-po.gc.ca/sci/protocol/shorekeepers/Guide/SKintro.pdf

Karr, J.R. d E.W. Chu. 1999. Restoring Life in Running Waters. Island Press, Washington, DC.: 206 pp.

Krebs, C.J. 1999. *Ecological methodology*. Benjamin/Cummings, Menlo Park, 620 p.

Milliken, G.A. and Johnson, D.E. (1984). Analysis of Messy Data: Volume 1. Van Nostrand Reinhold: New York.

Smiley, B. and C.D. Levings, 1996. The Shorekeepers' Guide to Surveying and Mapping Backshore and Intertidal Habitats, Plants and Animals: The British Columbia Edition. Background and Introduction. Prepared for the Oceans Science Program Fund, Fisheries and Oceans Canada, Pacific Region, Vancouver BC. 5 p.

Underwood, A, J. (1991). Beyond BACI: Environmental design for detecting human environmental impacts on temporal variations in natural populations. Aust. J. Mar. Freshwater Res. 42: 569-587.

Table 1: List of surveyed sites, their respective Shorekeepers codes, survey dates, and code letter for survey leaders. Note: "not stated" means that the leader's name was not stated in field sheets.

Site code	Site name	Team leader	Survey date
SK1997001	Pipers Lagoon #1	leader A	30-6-1997
		leader B	24-6-1999
SK1997002	Pipers Lagoon #2	leader A	1-7-1997
		leader A	20-7-1998
		leader B	9-7-1999
SK1997003	McGuffie Road	leader A	2-7-1997
		leader A	21-7-1998
		leader B	26-5-1999
		leader B	11-7-1999
SK1997004	Entwhistle, Nanaimo	leader A	3-7-1997
		not stated	22-7-1998
		leader B	31-5-1999
		leader B	12-7-1999
SK1997005	Lavender Road, Lantzville	leader A	4-7-1997
		leader A	23-7-1998
		leader B	10-7-1999
SK1997006	Eby Road, Lantzville	leader A	5-7-1997
		leader B	4-9-1998
		leader B	27-7-1999
SK1997007	Huddlestone Road	leader A	6-7-1997
		leader A	24-7-1998
		leader B	13-7-1999
SK1997008	Tyee Crescent, Nanoose Bay	leader A	7-7-1997
		leader A &	25-7-1998
		leader B	25-6-1999
SK1997009	Seadog Road, Nanoose Bay	leader A	8-7-1997

		leader B	16-8-1998
		leader B	17-5-1999
		leader B	16-7-1999
SK1997010	Seahaven, Nanoose Bay	leader A	30-7-1997
		leader B	4-8-1998
		leader B	26-7-1999
SK1997011	Wall Beach	leader A	31-7-1997
		leader B	19-8-1998
		leader B	29-7-1999
SK1997012	Gervis Road Access	leader A	2-8-1997
		leader B	6-8-1998
		leader B	28-7-1999
SK1997013	Driftwood Beach Access	leader A	3-8-1997
		leader B	5-8-1998
		leader B	14-7-1999
		leader B	6-8-1999
SK1997014	Blue Heron Park	leader A	4-8-1997
		leader A	10-8-1998
		leader B	8-8-1999
SK1997015	Madrona Point, Nanoose Bay	not stated	5-8-1997
		leader B	11-8-1998
		leader B	7-8-1999
SK1997016	Dogwood Road	not stated	13-8-1997
SK1997017	Stalker Road, Gabriola Island	leader A	14-8-1997
		leader B	5-9-1998
		leader B	13-8-1999
SK1997018	Drumbeg Park, Gabriola Island	leader A	15-8-1997
		leader B	26-8-1999
SK1997019	Josef Point #1, Gabriola Island	not stated	16-8-1997
		leader A	9-8-1998

		leader B	11-8-1999
SK1997020	North of Drumbeg Park #1	not stated	17-8-1997
		leader B	8-8-1999
SK1997021	North of Drumbeg Park #2	not stated	18-8-1997
		leader A	27-8-1999
SK1997022	North of Drumbeg Park #3	not stated	19-8-1997
		leader A	8-8-1998
		leader B	24-9-1999
SK1997023	North of Drumbeg Park #4	not stated	20-8-1997
		leader B	8-9-1998
SK1997024	Taylor Bay, Gabriola Island	leader A	21-8-1997
SK1997025	Nanoose Harbour	not stated	11-9-1997
		leader B	23-8-1998
SK1997026	Brant Viewing Point, Qualicum Beach	not stated	12-9-1997
		leader B	17-8-1998
		leader B	23-8-1999
SK1997027	Surfside Drive	not stated	13-9-1997
		leader B	22-8-1998
SK1997028	Columbia Beach	not stated	14-9-1997
		leader A	3-9-1998
SK1997029	Arlette Road, Parksville	not stated	15-9-1997
		leader B	21-8-1998
		leader B	22-9-1999
SK1997030	Jacks Point Park	not stated	16-9-1997
		leader B	6-9-1998
		leader B	16-9-1999
SK1997031	South Winchelsea Island	not stated	17-9-1997
		leader B	7-9-1998
		leader B	25-8-1999
SK1997032	Gerald Island	not stated	18-9-1997

		leader B	24-8-1999
SK1997033	Josef Point #2, Gabriola	not stated	16-8-1997
		leader A	9-8-1998
		leader B	12-8-1999
SK1998001	Surfside, Parksville Beach	leader B	18-8-1998
SK1998002	Tyee Road (Gorge)	leader A	10-7-1998
SK1999001	Rathtrevor Beach	leader B	28-6-1999
SK1999002	Surfside R.V. Park	leader B	14-6-1999
SK1999003	Nelson Road	leader B	30-6-1999
SK1999004	Ruckle Provincial Park	leader B	12-6-1999
SK1999005	Surfside Road (Marshal's), Qualicum Beach	leader B	1-6-1999
SK1999006	Moorcroft Camp	leader B	29-6-1999
SK1999007	Queen Elizabeth Promenade	leader B	26-6-1999
SK1999008	Rogue Gallery, San Pareil	leader B	16-6-1999
SK1999009	Invermere Road	leader B	15-6-1999
SK1999012	Moses Point Beach	leader A	15-7-1999
SK1999013	Ten Ten #1	leader A	27-7-1999
SK1999014	Ten Ten #2	leader A	28-7-1999
SK1999015	Tseycum #1	leader A	29-7-1999
SK1999016	Towner Park #1	leader A	23-8-1999
SK1999017	Downey #1	leader A	26-8-1999
SK1999018	Towner Park #2	leader A	25-8-1999
SK1999019	Hagan Bight-Kennes	leader C	30-7-1999.

Table 2: Number of surveys in which a quantitative step was incorrectly done or reported (%, with respect to totals, in parenthesis). There is a discussion for each step in the text.

7.1.1.1	N	step 13	step 15	step 18	step 21	step 22	step 23	step 24	step 27	step 31
1997	33	20 (61%)	5 (5%)	3 (9%)	1 (3%)	7 (21%)	1 (3%)	10 (30%)	33 (100%)	31(94%)
1998	28	16 (57%)	8 (7%)	3 (11%)	2 (7%)	2 (7%)	1 (4%)	13 (47%)	28 (100%)	20(71%)
1999	48	28 (58%)	8 (7%)	1 (2%)	2 (4%)	8 (17%)	1 (2%)	17 (35%)	48 (100%)	33(69%)
TOTAL	109	64 (59%)	21 (19%)	7 (6%)	5 (5%)	17(16%)	3 (3%)	40 (37%)	109(100%)	84(77%)

Table 3: Descriptive statistics for differences between surveys 1 and 2 in habitat elevation, habitat slope, habitat area, and the number of taxa identified. Differences may be real, depending on how the habitats were identified and the boundaries observed.

7.1.1.2 Variable	N	Mean	S. D.	95% C. I.
elevation (% variation)	19	10%	27	-5% / 26%
slope (% variation)	14	-29%	88	-90% / 31%
area (% variation)	7	133%	257	-161% / 428%
area (absolute variation) (m^2)	7	367	695	-429 / 1162
% taxa identified on both dates	10	39%	21	24% / 54%
% taxa identified only in date 1	10	25%	11	17% / 34%
% taxa identified only in date 2	10	36%	25	18% / 54%

Table 4: Sites that were surveyed and audited in 2001.

Site	Date of first survey	Audit date
Moses Point #1 (SK1997012)	20 June 2001	28 August 2001
Ten Ten #2 (SK1999014)	21 June 2001	16 August 2001
Cadboro Bay (SK2001003)	4 July 2001	18 August 2001
Willows Beach (SK2001004)	5 July 2001	17 August 2001
Jimmy's Beach (SK2001013)	17 August 2001	19 August 2001
Esquimalt Lagoon - Colwood Creek (SK2001009)	15 September 2001	15 September 2001

Table 5. Average Variance Decomposition over all species

	Variance Component Decomposition			
Component	Mean %	Std %	Min %	Max %
Area	3	8	0	59
Year	3	7	0	66
Habitat type	3	8	0	47
Among Habitats of the same type	6	11	0	56
Quadrat	85	18	32	100

Table 6. Comparison of nominal quadrats surveyed and number actually present in database. Counts are number of habitat units. There is a discrepancy between two aspects of the database. Part of the database has a field for the number of quadrats that were supposed to be counted. This was compared to the number of quadrats actually appearing in the spreadsheet. In some cases these differ. The "missing quadrats" are likely ZEROs, i.e. say six quadrats were surveyed, but in two of the quadrats, there were no species present. Rather than entering a "0", the data likely just wasn't entered. Why the number of actual quadrats reported surveyed could exceed the number with actual survey data is unknown? It is likely a coding error in part of the form, or more time was available, and so more data was collected and the initial form entry was not subsequently corrected?

Actual quadrat number with data present in spreadsheet	Quadrats nominally surveyed, i.e. initially indicated as surveyed						
	2	3	6	8	9	10	12
1	.	.	8	.	.	.	.
2	.	.	13	.	.	.	.
3	.	**1**	21	2	1	.	.
4	.	.	37	2	.	.	.
5	.	.	45	2	4	.	.
6	.	.	**194**	2	2	.	.
7	.	.	4	5	5	.	.
8	.	.	2	**13**	12	.	.
9	.	.	2	.	**31**	1	.
10	.	.	1	.	.	.	.
11		.	.	.	.	.	1
12							

Table 7. Comparison of number of quadrats with nonzero counts for a species (Periwinkles, code =55) vs number of quadrats surveyed. This is similar to Table 6 except looking only at perwinkles. For example, there were 45 surveys (sum of the counts in the row labeled 9 in the left-most column) where 9 quadrats were nominally measured, but 3 of the surveys only had 1 quadrat record reporting perwinkles, i.e. the other 8 quadrats likely had 0 counts and were not included in the database. Both Tables 6 and 7 are trying to show the extent of the problem when "0" and "missing data" are not distinguished.

Number of quadrats surveyed	Number of quadrats with actual counts											
	1	2	3	4	5	6	7	8	9	10	11	12
3	.	1	.	.	.	.	.	.	.	.	.	.
6	32	32	32	41	52	74	2	1	.	.	.	.
8	.	1	2	7	5	.	4	6	.	.	.	.
9	3	1	6	7	2	4	6	7	9	.	.	.
10	.	.	1	.	.	.	.	.	.	.	.	.
12	.	.	.	.	.	.	2	.	2	2	.	1

Appendix I: A "sample" analysis of the periwinkle data after inserting zeroes

(1) Insertion of zeroes:

For every habitat unit within a survey area, a comparison was made of the number of quadrats with data present to the number of quadrats nominally surveyed. If some quadrats appeared to be missing, they were inserted with counts of 0.

(2) Transform

Preliminary plots showed that the variation among quadrats increased with the mean. Not surprisingly, the standard deviation appeared to be proportion to the mean density indicating a Poisson process. A square root transform (sdensity = sqrt(density)) was used to try and make the variance approximately independent of the mean.

(3) The model fit was:

```
proc mixed data=expand;
  title2 'analysis on log scale - zeroes added - no interaction between year and habitat type';
  where speciesid='55';
  by speciesid ;
  class year habtype habitatid area;
  model sdensity = year habtype  / ddfm=satterth;
  random area habitatid(area) ;
  lsmeans year;
  lsmean habtype;
```

The model includes fixed effects of year and habitat type (cobble, rock, etc), and random effects of area and of habitat units within each area. Because not all habitat types are measured every year, it is not possible to fit an interaction between the two effects. Because of the large component of variation at the quadrat level, no effects for the autocorrelation within habitat units over time was added to the model.

(4) Results:

The ANOVA table indicated changes over time and over habitat types

Effect	F Value	Pr > F
Year	4.54	0.0039
habtype	5.55	<.0001

Estimates of the marginal means (on the square root scale) are:

Year	Estimate	Std Error
1997	14.3	1.7
1998	10.7	1.7
1999	9.8	1.5
2000	8.1	1.8

For example, the average density (over all habitat types) is about 14.3^2=204 in 1997 while declining to about 8.1^2=65 in 2000, a 204/65=3 fold decline.

These results are for illustration purposes only and should not be taken as a definitive analysis of the data because of the problems noted earlier.

Appendix II: Summary plots of habitat data that had a substantial number of periwinkles. There are two plots for each habitat type corresponding to plots on both the original data scale and on a log scale. The last two plots in the attachment show results pooled overall habitat types. Such plots are for illustration purposes only and should not be taken as a definitive analysis of the data because of the problems noted earlier. Locations given are those used from Table 1.

Figure Aa. Periwinkle (#55) abundance on cobble (C).

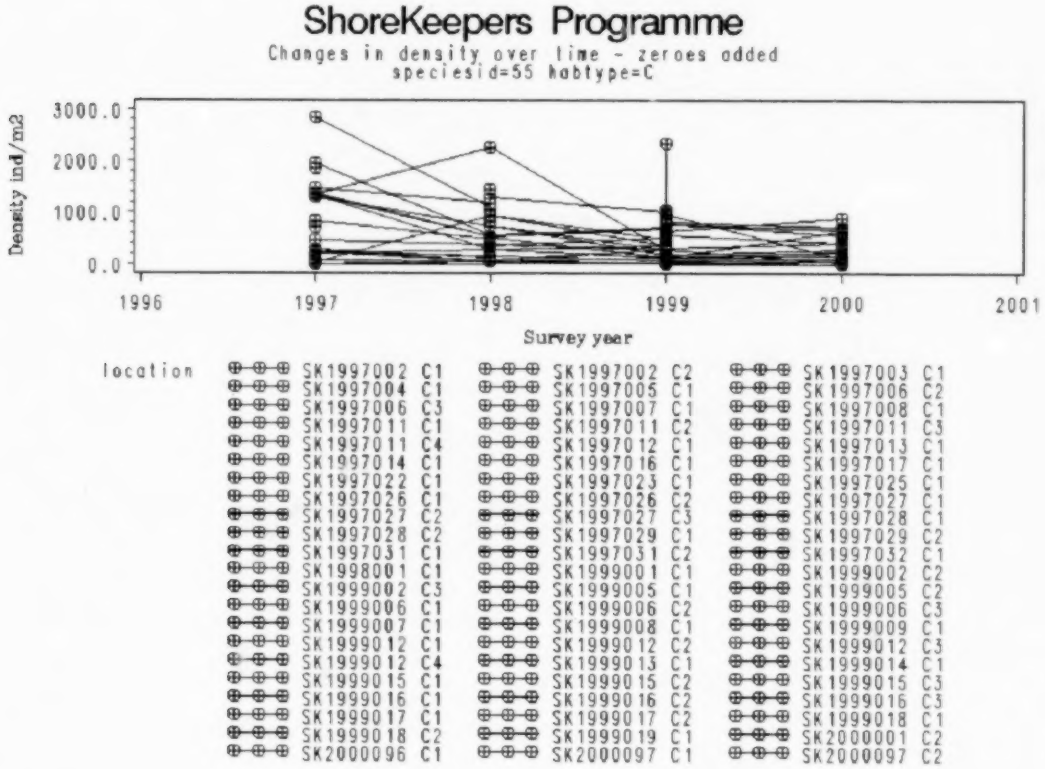

Figure Ab. Periwinkle (#55) abundance on cobble (C).

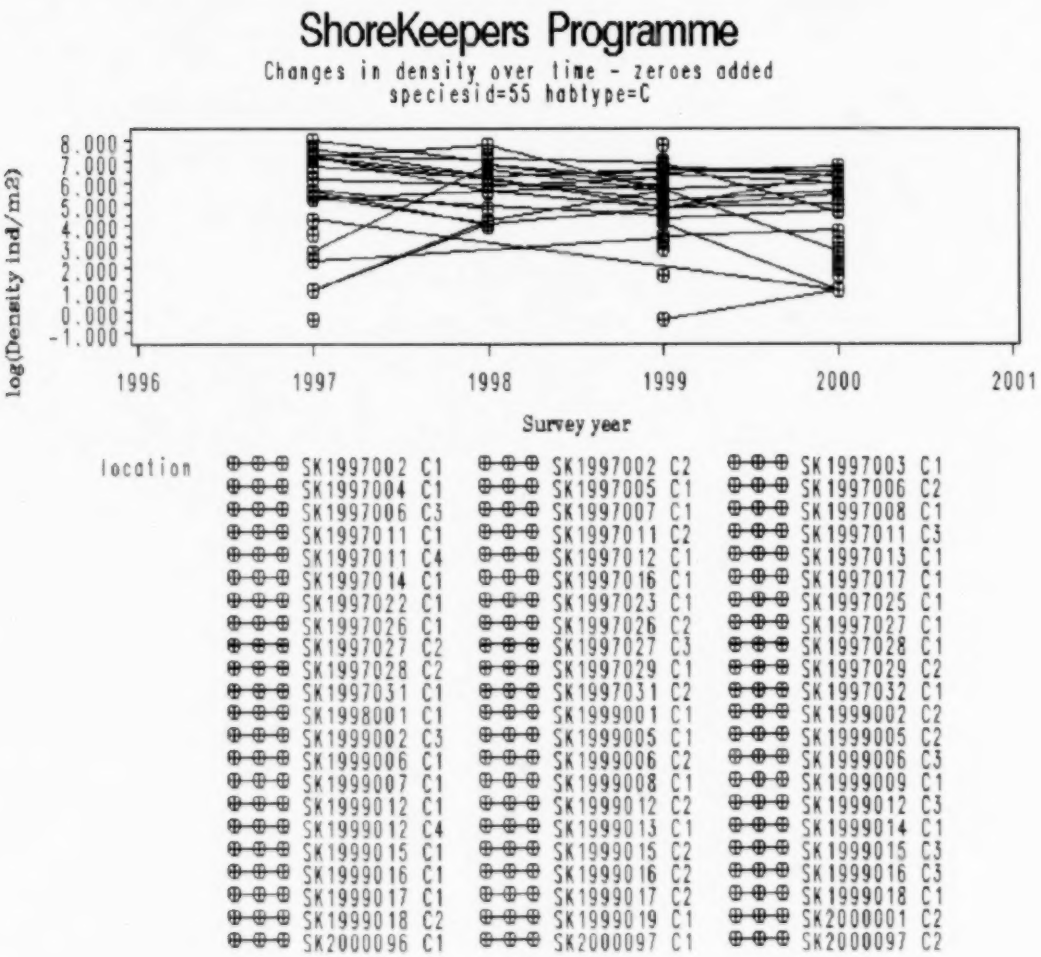

Figure Ba. Periwinkle (#55) abundance on *Fucus* (F).

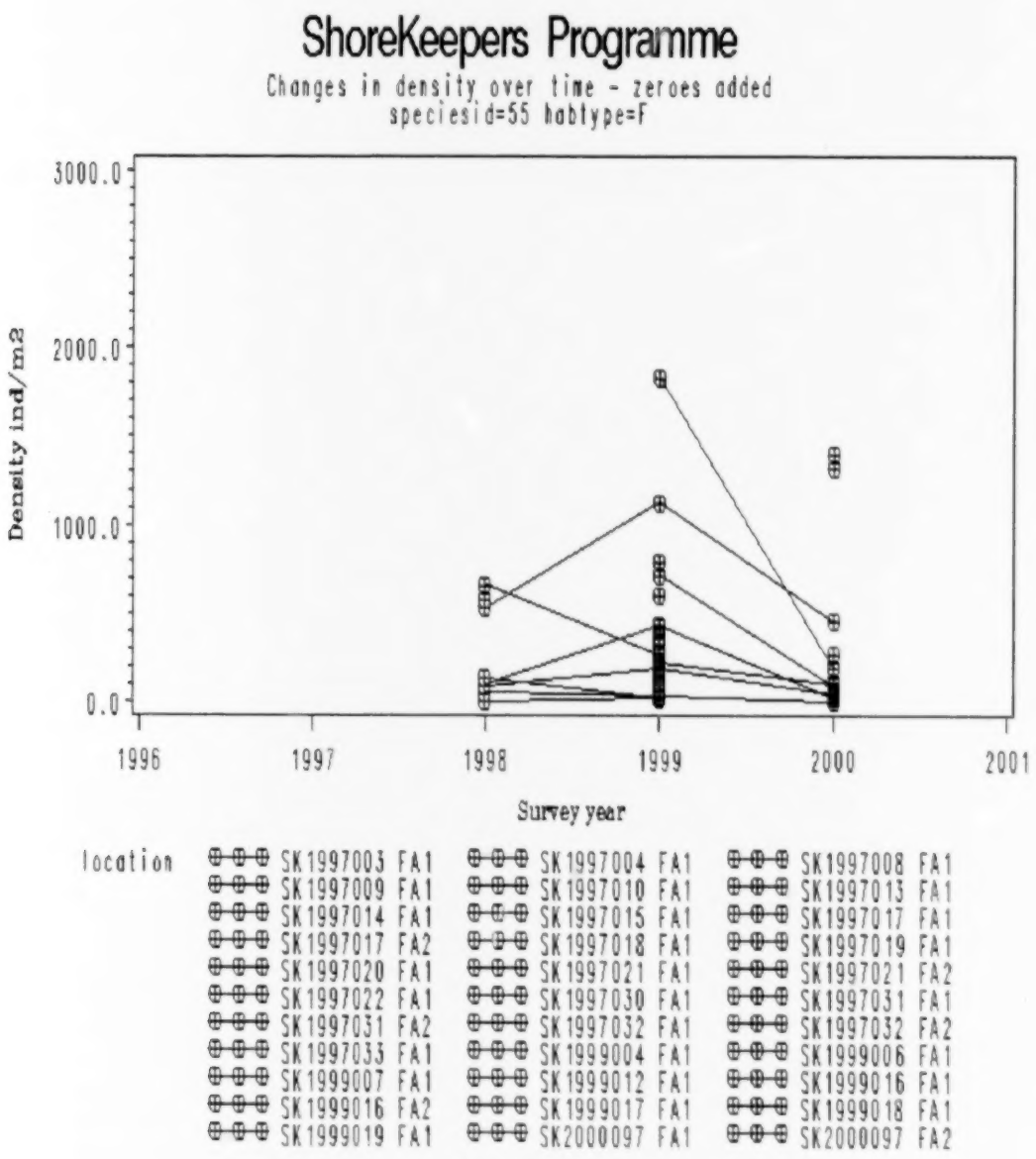

Figure Bb. Periwinkle (#55) abundance on *Fucus* (F).

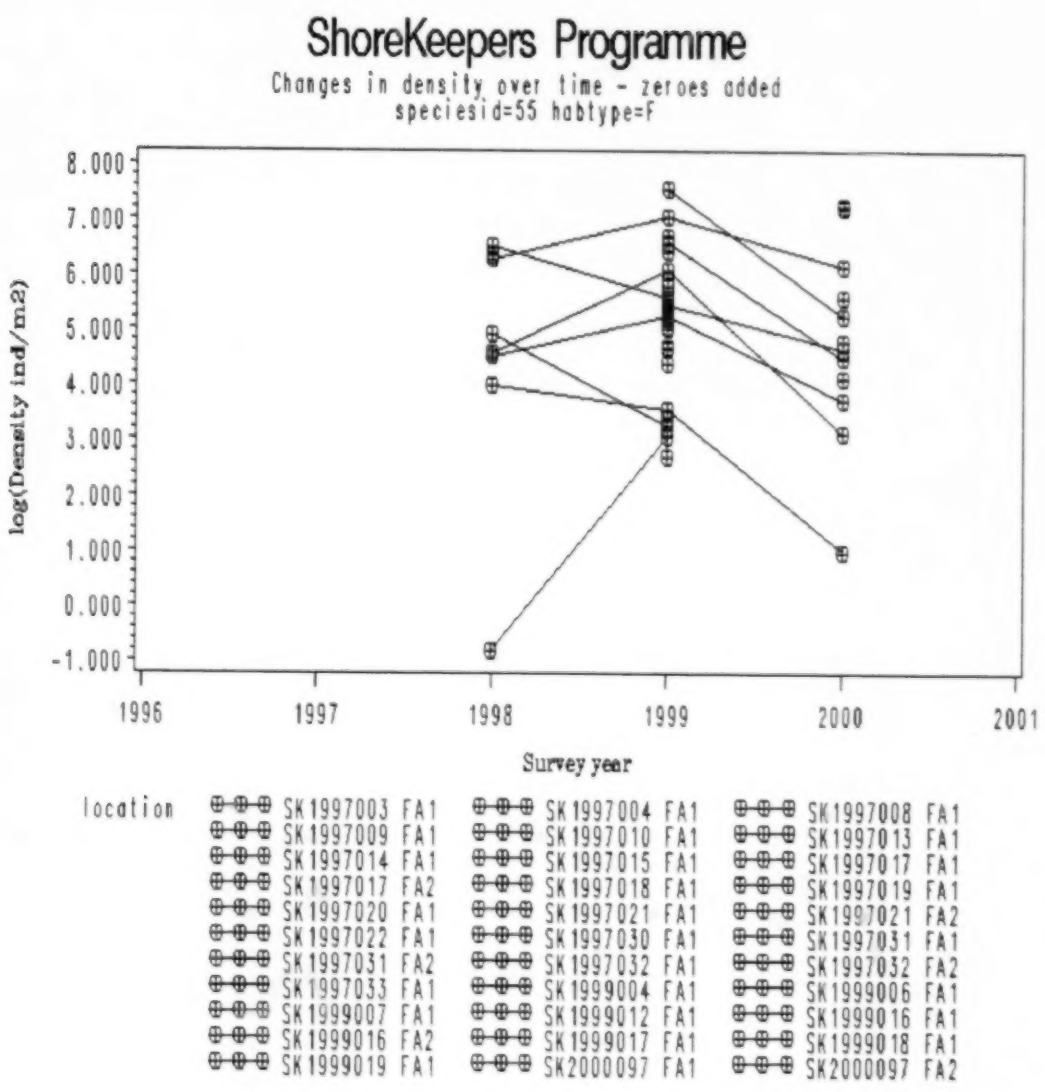

Figure Ca. Periwinkle (#55) abundance on "other algae" (OA).

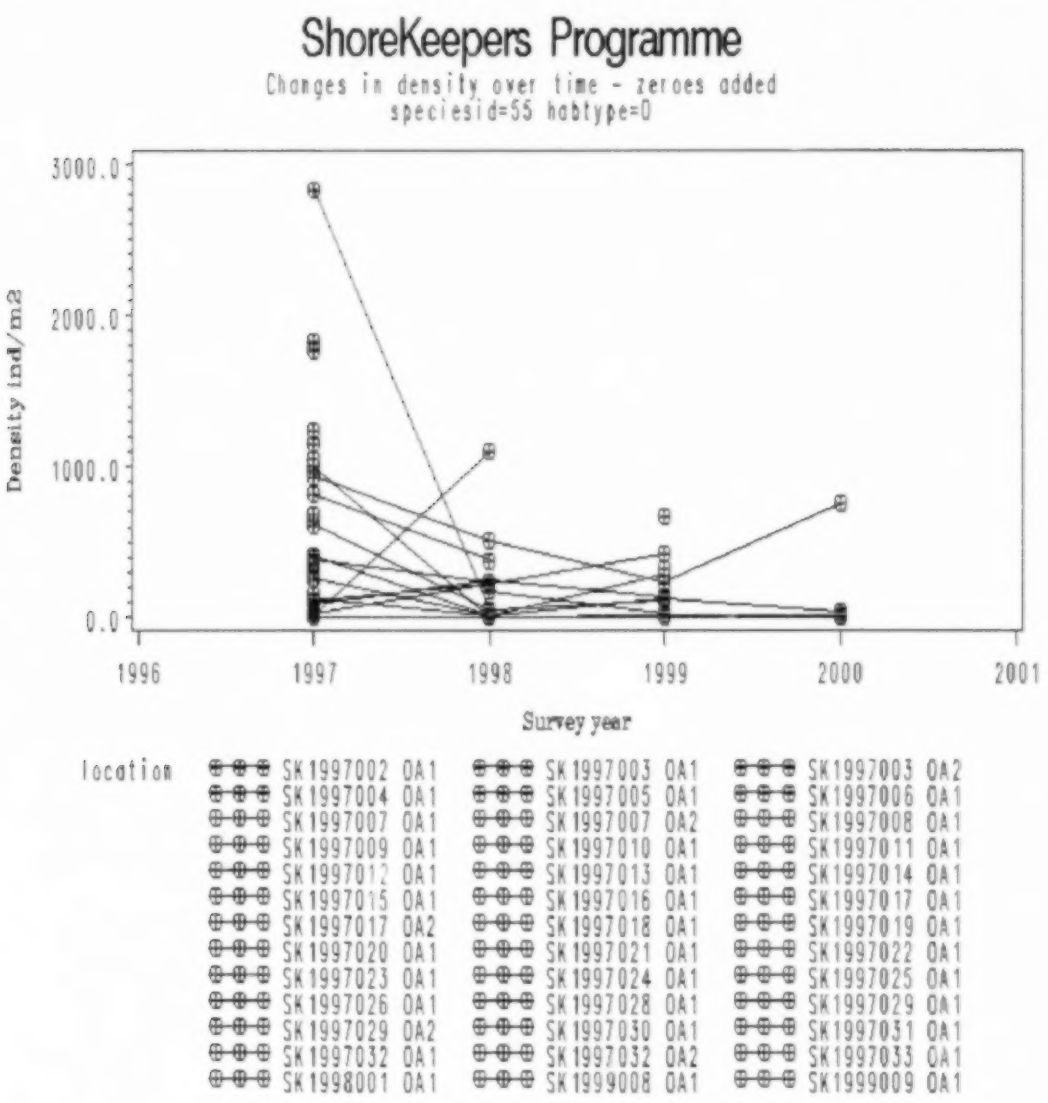

Figure Cb. Periwinkle (#55) abundance on "other algae" (OA).

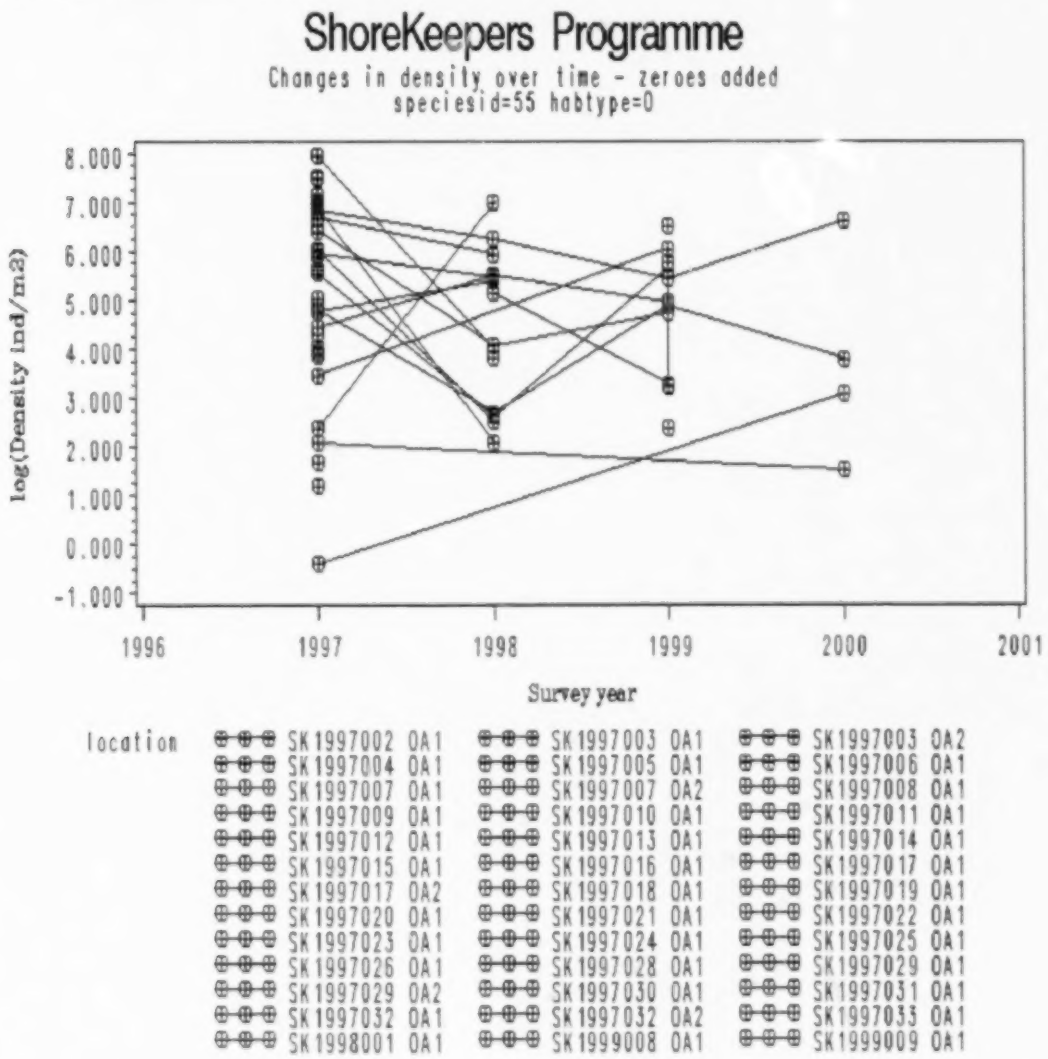

Figure Da. Periwinkle (#55) abundance on "rock" (R).

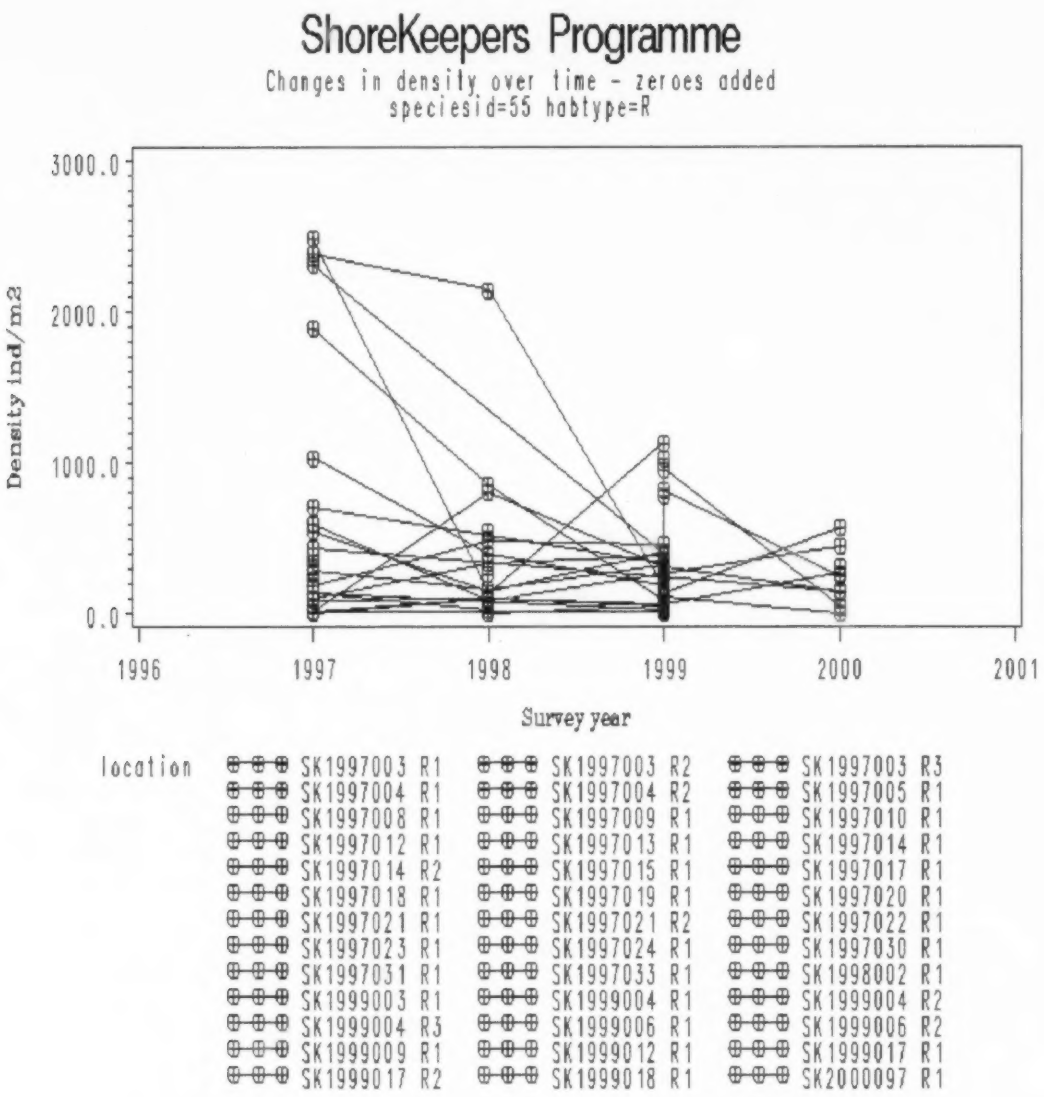

Figure Db. Periwinkle (#55) abundance on "rock" (R).

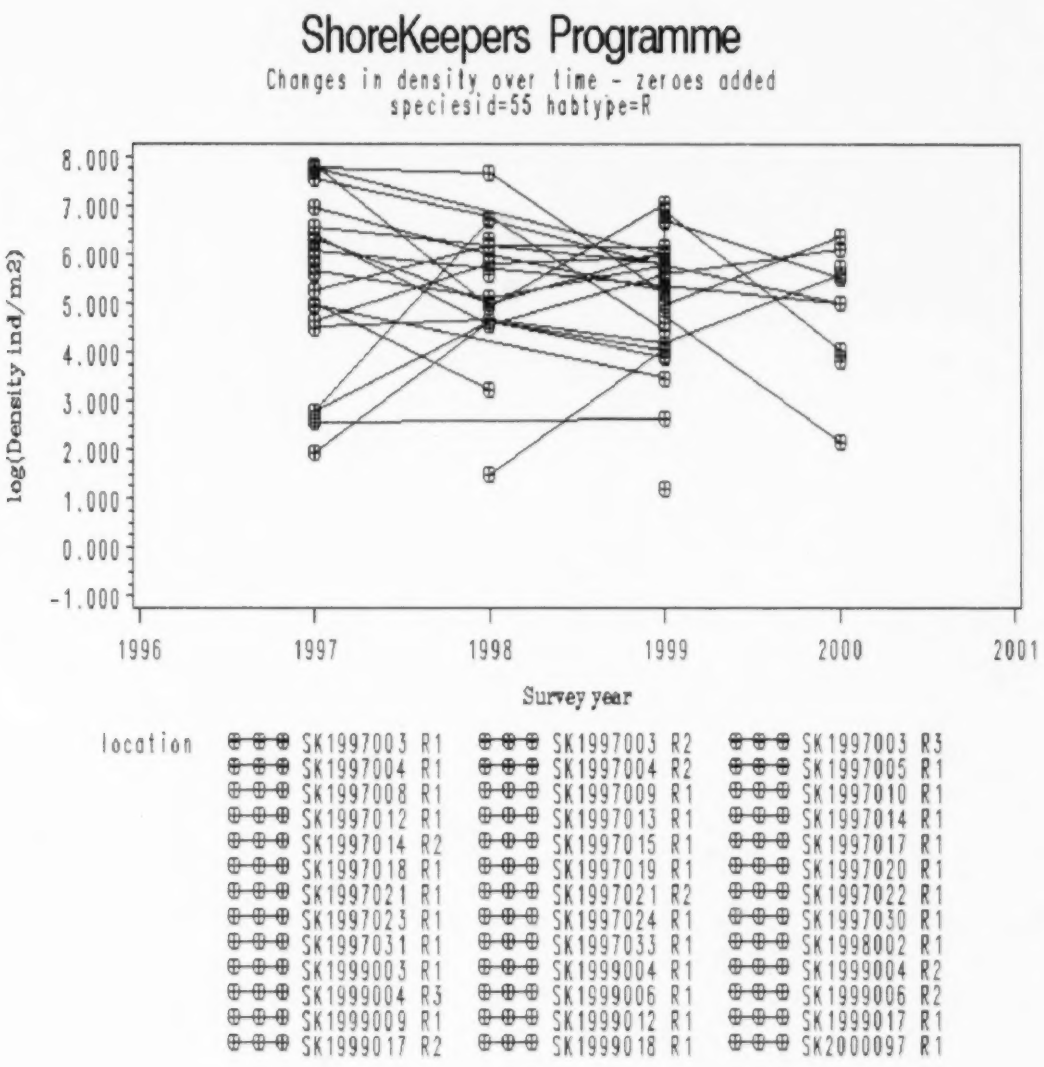

Figure Ea. Periwinkle (#55) abundance on all habitat types.

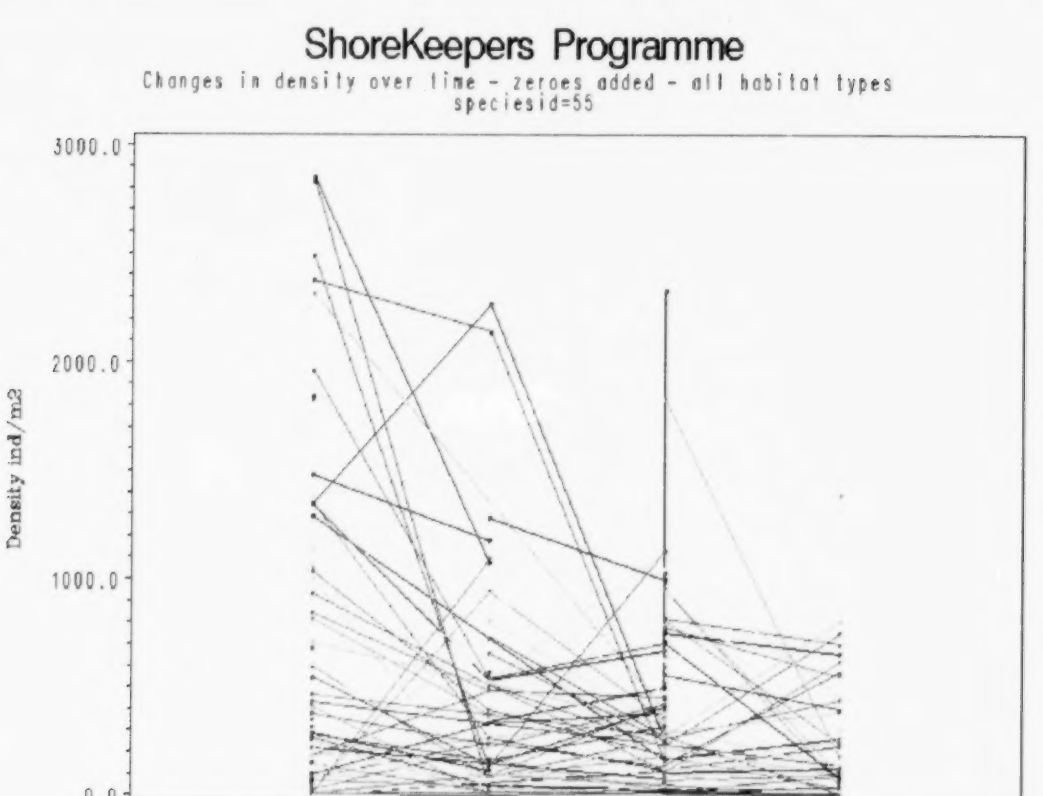

Figure Eb. Periwinkle (#55) abundance on all habitat types.